Mathematical Thinking
in the Measurement of Behavior

A Study of the Behavioral Models Project,
Bureau of Applied Social Research,
Columbia University

Mathematical Thinking in the Measurement of Behavior

Small Groups, Utility, Factor Analysis

EDITED BY *Herbert Solomon*

WITH CONTRIBUTIONS BY

James S. Coleman, Ernest W. Adams, Herbert Solomon

The Free Press of Glencoe, Illinois

Printed in the United States of America

DESIGNED BY SIDNEY SOLOMON

Library of Congress Catalogue Card No. 58—6488

LITHOGRAPHED FOR THE PUBLISHER BY
PROMOTION PRINTING & LITHOGRAPHING CORP., N. Y.

To Paul F. Lazarsfeld

Preface

Columbia University possesses a long tradition in studies which feature the interplay of mathematics and social science Particularly since the war, there has been much activity in this area in diverse parts of the University involving research in and teaching of mathematical methods for the social sciences, and the resolution of substantive problems in social science settings through mathematical approaches This activity has been integrated in formally at Columbia University by the accepted devices of department courses, faculty seminars and conferences, and sponsored research projects A more formal organized program, especially in connection with sponsored research projects, has existed in the Bureau of Applied Social Research at Columbia University since 1952

This volume is a product of the Bureau program and an outgrowth of the Behavioral Models Project, a project in the Bureau program sponsored by the Office of Naval Research from 1952—1956 It is the second of three volumes to appear from the work of this project The first volume entitled *Games and Decisions*, was published by John Wiley and Sons, Inc , New York in 1957 under the joint authorship of R Duncan Luce and Howard Raiffa The third volume, to be published simultaneously with or shortly after this volume by the Free Press, is edited by R Duncan Luce Professor Luce, now at the University of Pennsylvania, was previously at Columbia University where he played a central role in the mathematics program of the Bureau of Applied Social Research Professor Raiffa, now at Harvard University, and previously at Columbia University, also played an important role in the Bureau program The editor for this volume, who served as Mathematics Program Director for the Bureau, is now at Stanford University and his two collaborators also are now gone from Columbia University, James Coleman is at Johns Hopkins University and Ernest Adams is at the University of California in Berkeley In a way, then, Columbia has produced progeny for other centers

These three volumes are composed of analytical surveys of the uses of mathematics in several social science disciplines Some of these surveys were completed as long as three or four years ago but publishing difficulties delayed their appearance Thus they may suffer from some sins of omission because of the recent impetus in research in these subjects Emphasis in the surveys is on the uses of mathematical techniques and concepts which demonstrate the effectiveness of mathematical thinking in the resolution of substantive problems in social science No attempt has been made to inte

grate the subject matter of the surveys, either between volumes or within volumes

In planning the volumes, it was recognized that not all areas where mathematics has been successfully used in social science could be included However, it was hoped that by detailed and critical surveys in some specific areas the power of mathematical thinking in social sciences could be promoted to the extent that (1) mathematical and social science curricula would jointly undergo necessary revisions and that (2) social science research would become more of a science than an art

The topics discussed in the present volume are the choices of the authors and the editor The opening chapter by James Coleman on small group behavior and the following chapter by Ernest Adams on individual choice behavior represent areas of social science where the uses of mathematics are quite recent The final chapter, by the editor, on factor analysis, represents a topic which has been undergoing mathematical therapy for half a century This may give one explanation for the relatively small length of the last chapter as compared with the two preceding chapters Other explanations are too painful to contemplate

It is clear that this volume is not a text book in the ordinary sense However, it should be quite useful in courses that are currently being developed which feature the applications of mathematics in social science It has been used successfully in a graduate course at Columbia University on mathematical models in the social sciences It should be useful also to the many research workers in social science whose formal training has been of the non quantitative type, and to mathematicians and statisticians who have an enthusiasm for applications in social science and the development of techniques for social science problems

I would like to thank my Columbia colleagues — Professor Paul F Lazarsfeld, for providing the bubbling intellectual fervor and drive which made the whole program possible, and Professors T W Anderson, E Nagel, H Raiffa and W Vickery, for their help and counsel in guiding the program through the completion of the three volumes Many thanks are extended to R Duncan Luce for the strong and wise guidance he gave to research performed for the Behavioral Models Project and for his own stimulating research, both of which made the three volumes possible I would be more than remiss in my duty if at this point I did not cite the Office of Naval Research for its understanding, general helpfulness, and generous financial support of the Project from 1952—1956

HERBERT SOLOMON
Stanford, California
September 1, 1959

Contents — An Overview

PART ONE

The Mathematical Study of Small Groups

By JAMES S. COLEMAN

JOHNS HOPKINS UNIVERSITY

Contents

III. CONCLUSION

Prefatory Remarks(*)

THIS SURVEY consists of an intensive examination of four kinds of uses of mathematics in the study of small groups These four approaches are not the only uses of mathematics in this area, no attempt at exhaustiveness has been made On the other hand, they do represent distinctly different approaches, and thus give some idea of the different ways in which the tools of mathematics have been used to aid the study of small groups

There are three parts to the survey The first and shortest of the three is a brief attempt to locate mathematical models of small group behavior in the context of small group studies generally This is not an overview of research and theory in the area of small groups, but only an attempt to place the work to be examined in the larger framework of theory and research which includes non mathematical work The second, which constitutes the main body of the survey, examines each of four uses of mathematics in detail The third is a comparison and evaluation of the four approaches, suggesting the direction in which each is leading, and speculating about the kinds of results which may reasonably be expected from each

(*) I am grateful to Duncan Luce for discussion, comment, and criticism throughout the period of preparation of this paper, and to Anatol Rapoport and Herbert Simon for their comments on an earlier draft

INTRODUCTION

1. MATHEMATICS AND SMALL GROUP RESEARCH

IN RECENT YEARS, there has emerged a focus of research and theory in social science with the "small group" as its center. This work has covered a wide range of interests, attempting to answer such questions as:

Is the group a better problem-solving entity than the individuals who compose it, taken separately?

What is the effect of the group on the opinions of individuals within it?

What makes a group more or less cohesive?

How does the communication structure in a group affect its performance in carrying out tasks?

These and other questions have occupied the attention of numerous social psychologists and sociologists over the past fifteen or twenty years, and some answers have been forthcoming as a consequence of their work.

Systematic attention for small groups began in the early part of this century. One of the first men to emphasize the importance of small, face-to-face groups in modern social science was Charles Horton Cooley ([1925], p.23-28), whose concept of "primary group" is widely used even today. Cooley saw one of the major problems of society to be the socialization of the young the transformation into civilized beings of the infant "savages" who invade society by being born into it (*) Cooley suggested that certain kinds of associations with other people, that is, intimate, face-to-face, continuing associations, were the means by which such socialization takes

(*) Although Cooley was not the first to pose this as a problem, most social theorists before his time had taken it for granted, and not raised the question The work which developed from Cooley's beginning is a good example of the advance in a science which comes not from giving answers but from raising questions, seeing as problems things which before were simply taken as "obvious" This is a step in the building of a science which is too often overlooked in the search for answers to already-posed questions The *change* in a question, that is, perceiving the problem in a new light or posing a new question can often pave the way to solving problems which were previously unresolvable Conant [1950] gives good examples of this in his *Harvard Case Histories in Experimental Science* See "Robert Boyle's Experiments in Pneumatics," p 14, where Conant indicates the effect of raising a question about the similarity of the atmosphere to a "sea"

place He saw these 'primary group" relationships as a fundamental building block in society, serving as the basis for more elaborate forms of social organization

Despite Cooley's important contribution toward small group theory, however, the major work in this direction has been relatively recent, within the past twenty years Beginning in the late 1939's an awakening of interest in social psychology, and with it, in the study of small group processes, came about Today there exists a voluminous literature in the field

Small group behavior has become an important focus of social and psychological research

The Role of Mathematics But the use of mathematics in the study of small groups has followed some steps behind the resurgence of interest in small groups themselves This is to be expected, for mathematics is often used to formalize theories or laws which have been previously developed through experimentation and field work In any science, it ordinarily follows after empirical investigations However, in this branch of social science it seems that mathematical developments have lagged even farther behind empirical work than one might expect This is particularly surprising in view of the numbers of mathematically inclined social scientists who have become interested in developing mathematical models of small group behavior Workers at Harvard, MIT, Columbia, Michigan, and at many other centers have approached the study of small groups with the intention of developing mathematical models to describe their behavior, but there exist only a few such models in published form as a consequence of these intentions

This state of affairs immediately raises the question Why has there been no real proliferation of mathematical models? The answers to this question—for there appear to be several partial answers—lie in diverse areas To see what these different answers are, it will be helpful to examine closely the characteristic goals of investigations in small group behavior

The Goals of Research· Showing that a Relationship Exists Probably the most common goal of investigation, both field and experimental, has been to demonstrate the importance of one factor in affecting one or more other factors The Lippitt 'leadership climates' study [1940], for example, demonstrated the importance of the kind of leadership—democratic, authoritarian, or laissez faire—for group functioning (*i e* , for morale,

productivity, cohesion, etc.) Similarly, Muzafir Sherif's [1935] studies of the establishment of a group norm showed the importance of other persons' judgments for the judgments of each group member in an unclear situation. Even more conclusively, Asch [1951] showed, with his 'unequal line' experiments, the effect of majority judgments in distorting the judgments of individual group members. Lewin [1943], comparing the efficacy of a lecture and a group discussion in changing housewives' opinions about food, demonstrated the importance of group discussion in abrogating old norms and thereby instituting change. Carter [1949], Hemphill [1950], and numerous others have demonstrated the importance of certain factors—personality and otherwise—for leadership in a group. Roethlisberger and Dickson [1939] showed the importance of informal relationships for individual productivity and morale in work groups. The list could be extended indefinitely.

Many of these programs of research have been eminently successful as judged by the goals of the investigators. That is, they have demonstrated to the satisfaction of nearly all concerned that a particular factor had an important effect on some other variable, but they almost invariably stopped at that point. The strength of the relation, the precise quantitative form, and even more fundamentally, the precise conceptualization of the 'factors' involved, have often been lacking.

This is such a common occurrence in social science that we forget to be surprised by it. Yet other sciences have proceeded not simply by identifying the existence of a relation, but by studying the mathematical form of the relation, and by locating it carefully in a theoretical structure. Why this difference? Why have the aims of much investigation in small group research been limited to demonstrating that a relationship exists?

One reason seems to be the lack of agreement in this area on the variables or concepts which constitute the framework of investigation. Often in these studies the 'variables' which were shown to be related were not well enough defined to be amenable to precise measurement, nor did they have any place in a well developed theory. Such factors as 'authoritarian leadership' and 'democratic leadership' can be roughly identified, or at least contrasted when a comparison of groups is made. But to measure them quantitatively or to conceive of them as 'variables' or 'concepts' in a small group theory is quite another question.

However, even those variables which are easily measured have seldom been quantitatively related in any serious way. The variable of productivity, measured in terms of units of output or some similar quantitative measure, would seem to provide a good chance for developing quantitative relations

For example, one could establish a relation between units of output among a group of workers and the number of positive replies to questionnaires asking about feelings toward fellow members And, in fact, numerous studies have taken units of output by the group as the measure of their dependent variable But such quantitative measurements have ordinarily been used only to establish the existence of a relation, just as if the measurements were only qualitative comparisons It is easy to see why these measurements are tied completely to time and place and the particular conditions of observation, and could hardly serve to establish a useful quantitative generalization about group behavior In a similar way, many other factors which these investigations relate are merely crude indicators for the properties which are really of concern to the investigator This means, for example, that in an investigation of the relation between group cohesion and the amount of group discussion there is always the difficulty that no general measurement for amount of cohesion (or even for amount of discussion) exists apart from the specific conditions Measurements in such cases may serve the purpose of the specific experiment in showing that a relationship does exist, but they could not form the basis for a quantitative relationship of general value

One part of this problem is the fact that specific conditions surrounding the experiment always seem to influence the relationship under study These conditions are not easily specified, and their effect can be isolated only with difficulty Not knowing the effect of these various conditions, conditions which are not under study but which are nevertheless necessary in order to set up the experiment, there is little motivation to attempt quantitative generalizations Sherif, for example, in his experiments on the formation of group norms, knew that under different experimental conditions from those he used, a group norm would have been established at a different rate of speed and that the norm would have been of different strength He thus had to limit his results to the demonstration that, under certain specific conditions, a norm of some strength was formed within a certain length of time The difficulty in carrying such results further is again traceable to the problem of measurement and, more fundamentally, to the problem of concept formation Somehow, we have not yet found the concepts in terms of which quantitative theories can be developed, except in special cases, like the concept, 'size of group' If we had such concepts, then measurements under different experimental conditions could be compared They would not be so completely dependent on particular conditions of the situation

It might be said, in fact, that most social research in the area of small group behavior is still in the 'variable searching' stage The studies referred

to above, which have concerned themselves with demonstrating that a relationship exists, all seem to be engaged in the search for a fruitful way of looking at social-psychological phenomena. They are thus part of a pre-mathematical stage in research: a casting about, identifying the outlines of the phenomena to be studied, searching for the best concepts by which to represent the phenomena. When a science is at this almost pre-theoretical stage, it is foolish to ask that it clothe its casting about activities in mathematical form.

It appears, then, that investigators have been largely justified in eschewing quantitative relationships until a theory is developed which employs concepts easily measured in any situation and easily compared from one situation to another. At the same time, however, it might be that certain relationships have, like the Weber-Fechner law, a constant form over a wide range of situations (*). Then the use of quantitative relations would serve to show the constancy of this form, and would allow development of generalizations of somewhat broad scope. For example, Asch ([1951], p. 188) finds a certain relation between the amount of distortion of judgment and the size of the (unanimous) group majority against the subject. The amount of distortion increases up to a majority of four, then decreases slightly. Asch does not attempt to develop a quantitative generalization, presumably because the type of judgment and the experimental conditions so influence the results. However, suppose he could show that, as he varied the conditions, one parameter of the relationship changed (say the size of group at which the maximum occurred), but the shape of the curve which relates amount of distortion to size of the majority did not change. Then he would have a legitimate generalization, relating the group size at which maximum distortion of judgment occurs to the type of judgment required.

Such an approach may be one way to begin to use mathematics while small group research is still in its variable searching stage. More generally, there are a number of possible approaches to "partial mathematization" before agreement is reached on a theoretical framework which can be put in useful mathematical form. The work examined below will exemplify some of these approaches (†).

Emphasis on Single Relationships Apart from the concentration on the existence of a relationship, there is a second aspect of the goals of most

(*) The Weber Fechner law relates magnitude of discrimination to magnitude of stimulus and says roughly that the change in discrimination with a given increment in stimulus is inversely proportional to the existing level of stimulus.

(†) A discussion of various tactics and strategies which might provide footholds in the mathematization of social science may be found in Coleman [forthcoming]

small group research which may be a factor in inhibiting mathematical model-building. Besides being satisfied with demonstrating the existence of relations rather than their size or form, investigators have been concerned with *single isolated* relationships between two variables. In all the examples of work which were mentioned above, the investigators were concerned with single relations, that is, the effect of *A* on *B*. It is no coincidence that in the context of this emphasis on single relations, the above discussion was centered around single quantitative generalizations rather than mathematical models. The mathematical harvest from investigations of isolated single relations must be a harvest of single generalizations. Thus the deductive power of mathematics, which depends upon a network of *related* generalizations or propositions, is never used.

Mathematical models, at least those which are most evident in social science, have been models which can represent a dynamic system of behavior rather than isolated aspects of it. Or at least they have utilized some of the deductive power of mathematics. This emphasis, in conjunction with the emphasis of empirical work upon single isolated relations between two variables, has meant that there have been few meeting points between empirical research and mathematical models. Because of this incongruity, it may even yet be too early to look for a proliferation of empirically grounded mathematical models of small group behavior. For if these models require the conjunction of a number of related generalizations, then they must wait upon the development of these related generalizations.

Yet it is important not to overlook the value of isolated quantitative generalizations. In our eagerness to construct models of dynamic systems, we sometimes forget that mathematics is useful for expressing single relationships. In every science which is today quantitative, the introduction of mathematics did not begin with models of systems of behavior. It often began rather with these single relationships, transforming them first to precise quantitative form, and only then developing from them a theory or model of a system. Whether these generalizations were then "explained" by developing an underlying theory, or whether they themselves were treated as postulates from which a theory was synthesized, the theory or model was not the first step in introducing mathematics. (*) Keeping this in mind, it

(*) A good example of the difference between treating a generalization as a phenomenon for explanation or as a postulate for building a theory is provided by Boyle's Law (the pressure of a gas varies inversely with volume at constant temperature). Taken as a phenomenon to be explained, the molecular theory of perfectly elastic, *dimensionless bodies in constant* motion accounts for the law. But taken as a postulate in a theory itself, it is linked with Charles' law to form the perfect gas law, giving rise to deductions which neither Boyle's nor Charles' law could make alone.

may behoove us to pay more attention to this groundwork of mathematically expressed generalizations which can be based on isolated relations, and need not wait on a set of related propositions. It is true that such quantification fails to use the deductive aspects of mathematics, which is its greatest virtue. But even so, such quantified generalizations can stand as the building blocks from which deductive power can later issue, after the necessary related work has been filled in. (*)

Of course, some empirically grounded "models" do exist; those to be examined below all have some such grounding. And there is important and potentially very useful work being done with mathematical models which make no effort to be truly descriptive, but rather to show the implication of certain principles of behavior, *e.g.*, rational behavior, in complex situations. Nevertheless, the general point is probably true: the research emphasis on single isolated relations, together with the need for sets of interlocking relations if mathematical models are to be constructed, has inhibited the use of mathematics in this area.

The Study of Individual Behavior Disguised as Small Group Research. Another tendency inhibiting the marriage of research and mathematical models in small group behavior is of quite a different nature. Many studies in this area are in actual fact studies of individual behavior. The group is looked upon not as a system of behavior about which generalizations or theories are to be developed, but simply as a context within which the individual acts. The group serves as a proxy for the larger society, allowing the experimental study of influences otherwise not available to the laboratory. In other words, the "group" is often used as a manipulable social stimulus, and the focus of concern is upon behavior of an individual within it. Asch's studies of distortions in judgment, in which every member of the group but one acted on the instructions of the experimenter, is a good example of this use of the "group" (Asch [1951]). (†) Festinger's [1947] study of the expression of attitudes in a group situation is another such example. There have been hundreds of similar experiments. Almost all experiments in which "the group" is controlled by the experimenter and one individual is the naive subject are of the same sort, studying individual behavior in a group situation rather than group behavior as such. These experiments develop generalizations about the behavior of an individual under certain

(*) Although this discussion has emphasized *quantitative* generalizations, mathematically-expressed generalizations may be qualitative (for example, monotonic relations) as well. The first model to be examined in II is built of such generalizations.

(†) Nevertheless, Bernard Cohen (1958) has developed a stochastic model to account for the effect of the group's responses upon the individual's judgment.

social conditions, leaving unexamined the behavior of the social system which constitutes these conditions.

Such experiments as these can of course contribute to the development of models of group behavior but they can serve only as one component of such models. The complete model must deal with a system of behavior in which the individual's is only a part.(*)

Conclusion. In summary, at least three factors have served to inhibit the development of mathematical models for the behavior of small groups. These are (1) the tendency of research to demonstrate the *existence* of relations, rather than their precise form; this tendency is reinforced by the difficulties of measurement; (2) the tendency of research to focus on *single* isolated relations, rather than on functional interdependence, coupled with the disregard of mathematizing these single relations; (3) the tendency to use the small group in research as a context for studies of *individual* action, rather than as the focus of interest itself.

These three tendencies, and perhaps others, have inhibited the use of mathematics in this area of social science; but they have not completely stopped it, as the cases to be examined below indicate. These constitute diverse uses of mathematics in this area, each indicating one way to overcome some of the difficulties of theory- and model-building in this area. The first links together a set of qualitative generalizations about behavior in groups, and draws some deductions, again qualitative, from these generalizations. The second establishes a quantitative generalization about relative participation rates in small discussion groups, and attempts to develop a model to account for this generalization. The third focuses on the structure of relations within groups, developing ways for using mathematics to charaterize such structures. The fourth is work on models of "group action," that is, action of some unitary sort in a group. The four examinations, together with the comparisons and evaluations which follow them, may give some indication of fruitful directions for the use of mathematics in the area of small groups.

(*) It might seem that the generalization about an individual's behavior could be used *n* times (for a group of *n* individuals) to provide *n* simultaneous equations from which a model of the system of behavior can be constructed While this may often be possible *in principle*, it is hard to point to any instances in which this possibility has been practically feasible.

AN EXAMINATION OF FOUR KINDS OF MODELS

2 MODELS WHICH EXAMINE THE JOINT IMPLICATIONS OF QUALITATIVE RELATIONS

Introduction. The first case for examination is an approach which takes off from existing propositions in social research—the usual kind of qualitative propositions which say, in effect, 'As *A* increases, it produces an increase in *B*' The mathematization goes on to examine the joint implications of several such propositions taken together

Two examples of this kind of work exist in the literature of small group research The first, carried out by Herbert Simon [1952], was a formalization of some of the propositions stated by Homans in *The Human Group* [1950] This case will be examined in detail here The second example of such work is by Simon and Guetzkow [1955], and is based on work by Festinger and others This case will be only mentioned at the end of Section 2

It is useful to set down the general steps taken in a case of theory-construction or model building, to see just where the work begins and what its final aim is In the case of Homans' generalizations and Simon's formalization of them, the general procedure was something like this 1) observation and statement of relations limited to the time and place of the observations, 2) on the basis of a number of such specific relations, the development of generalizations cutting across these particular areas of content, 3) the formal restatement of these generalizations, in a form amenable to mathematical manipulation, 4) the drawing of deductions from these generalizations, which now constitute postulates of a system, 5) the comparing of such deductions with observation The diagram below indicates the general approach, it indicates as well which of the steps were taken by Homans and which by Simon

Homans took the published work of others who reported the behavior of particular groups and generalized from these to arrive at a number of

[observation and reporting] —(1)→ [empirical examples] —(2)→ [verbal generalizations or propositions] →

$\underbrace{\text{observation and reporting} \rightarrow \text{empirical examples}}_{\text{various original investigators}}$ $\quad$ $\underbrace{\text{verbal generalizations or propositions}}_{\text{Homans' reconceptualization}}$

(2a) → [restated generalizations] —(3)→ [mathematical postulates] —(4)→ [deductions from postulates] —(5)→ [observation of cases to verify the deductions and indirectly the postulates]

$\underbrace{\text{restated generalizations} \rightarrow \text{mathematical postulates} \rightarrow \text{deductions from postulates}}_{\text{Simon's model building}}$

generalizations much broader in scope than the original investigations. The generalizations relate three kinds of properties of a group or a relationship: sentiments toward others in the group, interaction with other group members, and the activity carried out by the group. Simon took some of these generalizations, stated them in mathematical form as postulates of a dynamic system (after first restating them in slightly different form — this is step (2a) above), and then drew deductions from them. The fifth step — comparing the deductions with observations—has not been carried out by anyone in any detail. The reasons for this will become evident later, when the qualitative nature of the deductions becomes apparent.

The present examination will focus primarily upon steps 3 and 4; it is these which constitute the kind of 'model building' which starts from existing social or psychological theory. Questions relating to steps 1 and 2 will be examined only under the heading of definition and measurement of the variables of the model, and these questions will be left until later, since Simon does not attempt to handle them in his development of the model (nor does Homans give precise definitions or measurement prescriptions in his development of the generalizations).

It is not hard to visualize the kind of groups Homans and Simon have in mind in developing the theory. Social clubs, work groups, athletic teams, are all representative of the kind of group to which these generalizations are meant to apply. The one empirical study Homans examines in detail before stating these generalizations is the Western Electric study of Roethlisberger and Dickson [1939]. This study, or the part of it which Homans uses, is a detailed observational study of an electrical relay wiring group. The men worked together the whole of every work day, and some saw each other socially after work. Other groups which Homans examined later in his book were quite different: a group of engineers in industry, a street

corner gang in Boston, a small primitive tribe. All were groups in their natural habitat, and not groups experimentally formed for the purposes of investigation.

The kind of generalizations Homans develops are something like this: If people come into sustained interaction, for whatever reason, sentiments of liking or friendship will grow up; these sentiments will increase both the number of activities they will engage in together, and the amount of interaction; these will in turn increase the friendliness, and so on, so that after a length of time the group is bound together by strong sentiments and by patterns of interaction and activity.

Generalizations of this sort have been current in social and psychological theory at least since sometime around 1900, when Cooley described substantially the same relationships. The main difference in Homans' work is the relative precision with which he states the relations (the rough paraphrasing above is not to be taken as a sample), and his explicit delineation of three types of variables: those concerning *interaction* between members of the group, those concerning the kind of group *activity*, and those concerning *sentiments* between members of the group (*).

The generalizations with which Simon works are not all those included in Homans' work, but only the three most fundamental ones. They relate the amount of interaction of the members ("I" in the equations to come), the "sentiments of liking" or 'friendliness' among them ('F' in the equations), the amount of activity actually carried on by the group ("A" in the equations), and finally, the amount of activity imposed on the group by the external environment ("E" in the equations). Of the four variables, the first three are variables whose value depends upon the values of the others at a preceding time ("endogenous" variables) while the last is not affected by the others (an 'exogenous' variable). That is, interaction, friendliness, and activity are all interdependent and are dependent upon E, that activity imposed by the group's environment.

Simon saw in these verbally stated interdependencies the possibility of setting up simultaneous equations from which deductions could be unambiguously drawn. It is important to note that *if* the verbal statements are made without ambiguity, and *if* the translation to the mathematical form is made correctly, there is no logical difference between the verbal statements and the mathematical ones. There are certainly psychological differences: most social scientists feel more at home with the verbal statements; however, most of us are unable to draw from them by mental mani-

(*) For a fuller statement of what Homans means by these terms, see ([1950] p. 34-38).

pulation the deductions which we can make from explicit equations through mathematical manipulation It is such differences as these — which have important consequences for the development of a science—that lead us to use one form of expression rather than the other (More will be said about these and other differing consequences—and it is by no means a case of black and white — in III)

In the previous paragraphs, the terms 'interaction,' 'friendliness," "amount of activity," and so on, as well as such symbols as I, F, etc , have been used rather loosely As in any theory, each of these terms plays a double role In the first place, they all stand for some kinds of events or states of being in actual groups When the symbols are words like "interaction or 'sentiments of liking," then their meaning is to some extent shared by all social scientists, and even by laymen That is, all who understand the English language have some idea of what Homans means when he says, "*If the frequency of interaction between two or more persons increases, the degree of their liking for one another will increase, and vice versa* It is true that this meaning may not be fully shared, each of us who reads this statement might go about testing it in different ways, each making certain observations on groups and saying, '*This* is what Homans means by 'frequency of interaction' and 'degree of liking '" Such differences simply mean that there is some ambiguity in knowing just what these symbols represent or, to put it differently, what their operational definition is But for the moment the point is that one role of these terms is to symbolize actual events or states of being The same is true of the more abbreviated symbols I and F, they are simply used to replace the words, and whenever one sees the letter F in these pages, he can substitute for it the term "friendliness

The second role of these symbols is as terms which relate to *one another* in particular ways in a sentence or an equation Here the relation is completely on the level of abstraction, not referring to actual groups in the real world at all This is most easily seen in a set of simultaneous equations, such as those which Simon sets up, relating I, F, and A One may carry out detailed examinations of such symbolic systems without ever referring to the actual events or states of being which constitute the meaning of the symbols

Homans never completely separates his discussion of abstract relations between symbols and his discussion of the relation of these symbols to actual events He largely assumes that we all know what is meant, *i e*, what kinds of observations are implied, when we talk about 'frequency of interaction" and "degree of liking " Simon similarly never treats this problem in detail,

concentrating most of his effort on developing the abstract system and tracing out its implications

The examination of Simon's model, then, will be in two parts first, a discussion of the abstract system of relations between symbols (whether verbal or mathematical ones) and second, a discussion of the meaning of these symbols, that is, their relation through definition and measurement to the world of actual events in groups of people Until the section on measurement, the relation of the model or theory to actual events will be disregarded

The Abstract Relations. The verbal postulates which Simon states (differing slightly from those originally stated by Homans ([1950], p 112, 118), are these

(1) The intensity of interaction depends upon, and increases with, the level of friendliness and the amount of activity carried on within the group We will postulate, further, that the level of interaction adjusts itself rapidly—almost instantaneously to the two variables on which it depends

(2) The level of group friendliness will increase if the actual level of interaction *is higher than that 'appropriate' to the existing level of friendliness* That is, if persons in a group with little friendliness are induced to interact a great deal, the friendliness will grow, while if persons with a great deal of friendliness interact seldom, the friendliness will weaken We will postulate that the adjustment of friendliness to the level of interaction requires time to be consummated

(3) The amount of activity carried on by the group will tend to increase if the actual level of friendliness is higher than that appropriate' to the existing amount of activity, and *if the amount of activity imposed externally* on the group is higher than the existing amount of activity We will postulate that the adjustment of the activity level to the imposed activity level and to the actual level of friendliness both require time for their consummation

Simon derives two sets of equations from these postulates one set derives directly from the postulates, with no further assumptions about the forms of the equations, and the other makes certain linearity assumptions The latter set will be considered briefly later The former will be considered in more detail

From the first postulate is derived the equation

$$I = f(A, F) \qquad (2\ 1)$$

which is just a different way of saying, Interaction is a function of activity and friendliness" It is easy to see that the verbal postulate (1) certainly says more than this It says, in effect, that a *positive* change in A or in F produces a positive change in I We can express this mathematically by writing

$$\frac{\partial I}{\partial A} > 0 \text{ and } \frac{\partial I}{\partial F} > 0 \tag{2 1a}$$

The restrictions (2 1a) together with equation (2 1) constitute a formalization of the verbal proposition (1)

However, propositions (2) and (3) are somewhat more complicated Both these propositions say (a) that a level of the dependent variable exists "appropriate to" the level of the independent variables, (b) it takes some time for this level to be reached, and (c) if the independent variable is higher, the 'appropriate level of the dependent variable is higher Simon interprets (a) to mean that for a given level of interaction (I) (in proposition 2) there is an *equilibrium* level of friendliness (F) such that if everything else in the system were held constant, F would reach this equilibrium level This certainly seems to be a reasonable interpretation of (a) Simon interprets (b) to indicate that the equation relating I and F (again in proposition 2) should not describe only the final *state* of the system, but also the *rate* at which the state is approached, that is, a differential equation in F and time is implied

The equations implied by propositions (2) and (3) are

$$\frac{dF}{dt} = g(I,F) \tag{2 2}$$

$$\frac{dA}{dt} = \psi(A,F,E) \tag{2 3}$$

But again these formalizations neglect part of the substance of the two propositions Condition (c) above says [for proposition (2)] that when $\frac{dF}{dt}$ is zero, that is, when the level of F is appropriate to" what has been the level of I, an increase in I will raise the 'appropriate' level of F That is, it will increase F through making $\frac{dF}{dt}$ positive Stated simply, $\frac{dF}{dt}$ varies directly (in the neighborhood of $\frac{dF}{dt} = 0$) with I, or $\frac{\partial g}{\partial I} > 0 \left(\text{where } g = \frac{dF}{dt}\right)$ But even beyond this, in order that there exist an equilibrium value of F, it is necessary that for large values of F, as F increases $\frac{dF}{dt}$ must decrease (in the neighborhood of $\frac{dF}{dt} = 0$) Otherwise, the increase in I, producing a positive $\frac{dF}{dt}$,

would increase F, which would in turn increase $\frac{dF}{dt}$ and thus increase F to infinity The mathematical statement of the restriction which prevents this is $\frac{\partial g}{\partial F} < 0$ (*) *Stated in full, the restrictions on equations* (2 2) and (2 3) implied by condition (c) are

$$\frac{\partial g}{\partial I} > 0, \frac{\partial g}{\partial F} < 0 \tag{2 2a}$$

$$\frac{\partial \psi}{\partial F} > 0, \frac{\partial \psi}{\partial E} > 0, \frac{\partial \psi}{\partial A} < 0 \tag{2 3a}$$

Simon does not make these conditions explicit (†) Since he treats this general model only in a subsidiary way to the linear one which he also presents, he makes the transition from words to symbols quite rapidly He by-passes the statement of these conditions, stating only the slope of the curves(**) which these propositions imply (††)

It is important to understand clearly the verbal postulates and their mathematical restatements Verbal propositions or generalizations of this type are prominent in present-day social research, and it is thus valuable to see what kinds of models can be developed using only such generalizations The Simon-Homans model illustrates this situation well

It is not precisely true that no assumptions are added in the mathematical restatement of verbal propositions, as is carried out above There are

(*) This restriction, and the parallel one in (2 3a), are stronger than necessary, for the only necessary restriction is that these two inequalities hold for large values of F *But recognizing this, the restrictions can be made as in* (2 2a) and (2 3a), deferring until later the question of what difference it makes if they are weakened

(†) In the Simon Guetzkow model based on hypotheses by Leon Festinger, the translation is made explicit Some of the translations carried out there are more complex than these, and constitute good *case examples in the translation of a verbal theory to a mathematical one*. See Appendix 2 1, p 38 for a short discussion of this model

(**) *The curves mentioned are those which relate A and F when* $\frac{dF}{dt} = 0$ and when $\frac{dA}{dt} = 0$ They will be introduced when the deductions to be made from the postulates are discussed

(††) In doing this, he at one point deduces from the verbal proposition 3 only the fact *that the slope of each curve is positive* whereas he also needs to deduce a second condition as well, namely that $\frac{dA}{dt}$ is negative to the right of the curve, positive to the left (Simon [1952] p 208) Lacking this, the graphical deductions are not possible *It is likely that he merely neglected to add the requirement, the deductions following it proceed as if he had included it*

two assumptions which may or may not exist in the verbal theory, but which the mathematical reformulation helps make explicit. One is the assumption that the variables are "state" variables of the system, which have certain relations independent of the time path of the system. In this model, the assumption is that the equations (2.1), (2.2), and (2.3) between I, F, A, and E are independent of the length of time a group has been in existence or how it came into existence.

The second assumption concerns the nature of the relation assumed to exist between two variables when one says, "As X increases, Y will increase." It is possible to construe such a statement as implying a direct structural relationship, with Y dependent on X in the sense that if all other variables in the system remain constant, an increase in X will be followed by an increase in Y. On the other hand, it is possible to construe the statement as meaning that X and Y will change in the same direction when some other variable which affects them both is varied. If the statement is construed in the latter way, then it is not legitimate to posit an algebraic relation between the two variables and embed this in a system of relations as has been done in the model discussed here. If a mathematical model is to be built from Homans' statements, it is necessary to assume that the verbal statements are meant to refer to a direct structural relation, rather than simply to co-variation between the two variables.(*)

It is clear if one examines Homans' verbal generalizations and the textual material surrounding them, that he evidently meant the variables as variables of state, and the relations as direct structural relations. These assumptions which must be made explicit if a system of equations is to be set up will only be mentioned here. The important point to recognize is that such assumptions ordinarily exist in the verbal form of the theory, and one value of a formalization lies in making them explicit. The specific assumptions mentioned, that the variables are state variables and that the equations represent structural relations rather than mere co-variation, are two of the most important and most difficult to confirm of any that social theorists make.

Deductions from the Postulated Relations. All the above points concern the translation of the verbal generalizations (as stated by Simon) into the formal system of relations. These relations constitute the postulates of

(*) See Simon [1957] for a discussion of structural relations and co-variation. They there make explicit the assumption that the verbal statements of Festinger are meant to imply structural relations.

the formal model or theory, and the question is, what can be done with them? As with any postulates, their usefulness is found by drawing implications or deductions from them (*) If they are so weak that almost nothing can be deduced from them, they are not of much use, and if deductions can be drawn, but these disagree with observed data, then the postulates are of use in proving themselves incorrect Or if they contain an inner contradiction, so that deductions drawn from a certain subset of them are in conflict with those drawn from another subset, they are of use only in proving themselves contradictory In this last case, where there is internal contradiction, the inconsistency is shown without ever examining empirical data, the formal system is simply a logically inconsistent one It is likely that this would be the fate of many vaguely stated verbal theories in present day social science if they were held up to the cold glare of mathematical formalization In such cases, the real gain would not come from showing conclusively that the theories were inconsistent, it would come through showing just *where* and *how* the theories were inconsistent, thus directing energies toward their modification

These, then are some of the results which arise from carrying out deductions on a set of postulates But what are the deductions which Simon is able to make from these postulates It is certainly true that the postulates are weak they say only what variables are related, and in which direction one changes when another changes An example of a stronger equation than these would be one which tells the *structure* or *form* of the relations between these variables Simon's linear system (which is mentioned later) is one such form, another would be multiplicative relations (*e g*, $I = kAF$ in place of equation (2 1))

The deductions Simon makes are based upon knowledge of the *direction* the system will be tending when one or both of the differential equations of the system are equal to zero, that is, when one or more of the variables is not changing over time The deductions will not be reproduced in full, but only enough to show the general approach (†)

First, the three equations are reduced to two by collapsing equations (2 1) and (2 2) Substituting (2 1) in (2 2),

(*) There are other possible areas of usefulness which have nothing to do with the deductions from the postulates these are largely psychological functions for the investigator or student such as fixing in his mind exactly what he has found or assumed Important as these may be they will not be considered here but in III when possible good and bad results for small group theory and research which may come from various kinds of mathematical model building are considered

(†) Simon presents all his deductions on pages 207 210 of his paper (Simon [1952])

$$\frac{dF}{dt} = g(f(A, F), F) = \phi\ (A, F) \tag{2.4}$$

and since $\frac{\partial f}{\partial A} > 0$, and $\frac{\partial g}{\partial f} > 0$, then

$$\frac{\partial \phi}{\partial A} = \frac{\partial g}{\partial A} = \frac{\partial g}{\partial f} \cdot \frac{\partial f}{\partial A} > 0. \tag{2.4a}$$

This inequality, $\frac{\partial \phi}{\partial A} > 0$, supplies one of the restrictions on (2.4). The second, however, does not so directly follow. Since $\frac{\partial f}{\partial F} > 0$ and $\frac{\partial g}{\partial f} > 0$, then the effect of F working through f (that is, through I, since $I = f$) is to increase g. But the effect of F directly on g is, as stated in (1.2a), to *decrease g*. Thus the partial derivative $\frac{\partial \phi}{\partial F}$ in equation (1.4) may be positive or negative, depending upon whether the influence of F on g through f or its direct influence on g is greater. That is, the rate of change of friendliness may either increase or decrease as friendliness itself increases. The equations do not predict which will happen. But if $\frac{dF}{dt}$ kept increasing with an increase in F, then once friendliness started going up, it would continue in an ever-increasing ascent with no limit. In order for this not to occur, two conditions are necessary. First, the effect of F upon must be negative:

$$\frac{\partial \phi}{\partial F} < 0 \text{ (at least for large } F\text{).} \tag{2.4b}$$

But beyond this, the increment in F (which tends to decrease ϕ) necessary to counterbalance the effect of a given increment of A (which tends to increase ϕ) must become smaller as F increases. Intuitively, what this means is that if A increases, thus increasing F through making $\frac{dF}{dt}$ positive, the resulting increase in F will be more than enough to depress $\frac{dF}{dt}$ back to zero or negative. Stated analytically, this requirement is that

$$\frac{\partial^2 \phi}{\partial A^2} < 0 \text{ or } \frac{\partial^2 \phi}{\partial F^2} > 0 \text{ (in the neighborhood of } \phi = 0\text{).(*)} \tag{2.4c}$$

(*) Simon states this requirement graphically, that the curve $\phi = 0$ (in a graph like the one on p. 27) will be concave upward

An assumption similar to this is necessary for equation (2 3) Like friendliness, activity is assumed *not* to increase without limit, implying a restriction on equation (2 3) that the curve $\phi = 0$ is concave downward, or analytically,

$$\frac{\partial^2 \phi}{A^2} > 0 \text{ (in the neighborhood of } \psi = 0) \qquad (2\ 4d)$$

If this were not so, activity would continue to increase without limit, because the increasing friendliness would increase $\frac{dA}{dt}$ more than the increasing activity would decrease it

With this assumption of an upper limit, or a "saturation effect," on F and A, it becomes possible to make qualitative deductions about the behavior of the system with respect to equilibrium These deductions are made by setting both equations (2 3) and (2 4), which are now the two equations defining the system, equal to zero This gives two relations between A and F, one for $\phi = 0$ and one for $\psi = 0$ Each of these describes a curve in the A, F plane, as in Fig 2 1 and Fig 2 2 At one of the curves (ϕ=0), F is constant through time, since $\frac{dF}{dt}$ is zero, and at the other A is constant through time At the intersection of the lines is an equilibrium for the system, for both F and A are constant That is, at the points of intersection there are no forces on either A or F to move the system from this point

These curves cannot be precisely located in the A, F plane, because equations $\phi(A, F) = 0$ and $\psi(A, F, E) = 0$, together with the restrictions upon them, give too little information However, these restrictions (2 3a, 2 4a, 2 4b, 2 4c, and 2 4d) do determine the general shape of the two curves Consider the curve $\phi = 0$ The three restrictions on it are restated

$$\frac{\partial \phi}{\partial A} > 0$$

$$\frac{\partial \phi}{\partial F} < 0$$

$$\frac{\partial^2 \phi}{\partial F^2} > 0$$

The first restriction implies that ϕ increases as the system moves to the right in the A, F plane The second implies that ϕ decreases as the system moves up in the A, F plane Taken together, the two restrictions imply that

the region where ϕ is positive will be to the lower right, while the region where ϕ is negative will be up and to the left. Thus the line separating these two regions, that is, the line $\phi = 0$ will have a positive slope, as indicated in Fig. 1.1. The third restriction, that the second derivative of ϕ with respect to A is negative, implies that as Γ increases, a smaller and smaller increment, relative to the increment in A, is required to keep $\phi = 0$. This means that the curve $\phi = 0$ must be concave downward, as shown in Fig. 2.1.

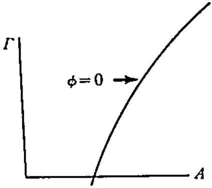


Fig. 2.1

The restrictions (2.3a) and (2.4d) on ψ are just the same as these, except that the axes are interchanged. Thus they imply that the curve $\psi = 0$ is shaped as in Fig. 2.2.

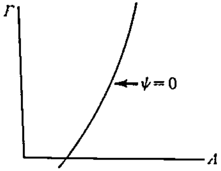


Fig. 2.2

These restrictions exhaust the information contained in the original postulates. What can be deduced from them about the behavior of the group through time? On the surface, very little, for the location of these curves relative to one another remains unknown. But upon assuming alternative locations, the alternative possibilities can be examined. One is represented in Fig. 2.3.

If the two lines have this relative location, the upper point, A_2, F_2, represents a point of stable equilibrium, while the lower point represents a

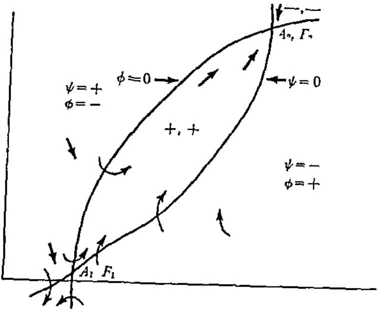


Fig 2 3

point of unstable equilibrium These results and others are arrived at by using the following information when the system (represented by any one of the arrows in the figure) is in the neighborhood of the curve $\psi = 0$, it will move in a vertical direction, and when it is in the neighborhood of the curve $\phi = 0$, it will move in a horizontal direction In each of the different regions of the plane, the system tends in the direction indicated by the arrows For example, when $\psi = -$ and $\phi = +$, the system is moving up and to the left Through such considerations, deductions may be made The remaining deductions will not be carried out here, but only stated They are in the form of propositions or theorems resulting from the kind of analysis begun in Fig 2 3

Before stating the propositions, it may be noted parenthetically that the empirical restriction of "saturation" upon propositions (2) and (3) does not precisely imply the restrictions which Simon (and we) have placed upon equations (2 2) and (2 3), that is, the restrictions of concavity (or stated analytically, inequalities 2 4c and 2 4d) That is, it implies concavity upward for $\psi = 0$ and downward for $\phi = 0$ *only* at extreme positive values of A and F This weaker empirical restriction opens up added possibilities

Graphically the system might look like this Here there would be three equilibrium points, with the upper (A_3, F_3) and lower (A_1, F_1) stable, and the center (A_2, F_2) unstable Groups which behaved this way would have two stable levels of activity and friendship, one with a relatively high

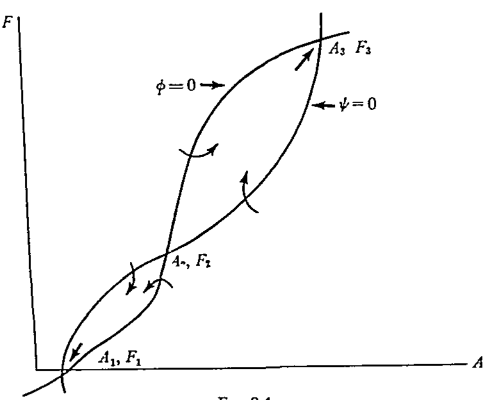


Fig 2 4

degree of activity and intimacy, and the other less intimate The levels of activity and friendship at which the group started would determine whether it would increase toward greater intimacy or decrease to a less intimate and active level Such a system seems reasonable on the basis of general experience

Though this possibility (and others, of course) exists, we shall not consider it further, but take as given the stronger restriction implied by Figs 2 1 — 2 3 The propositions Simon deduces are ([1952], p 209)

(1) First of all, there is at most only one point of stable equilibrium, toward which the system may tend If this stable equilibrium point exists, there may be a second point of unstable equilibrium. Like a pyramid balanced on its point, the group at this equilibrium point would fall over either on the side of increasing activity, interaction, and friendliness, or on the side of decreasing activity, interaction, and friendliness If both of these points (which correspond to points A_2, F_2 and A_1, F_1 in Fig 2 3) exist, then if the initial activity and friendship are high enough, the group will go toward the stable equilibrium If they are below a certain point, friendship and activity will progressively decrease until the activity is zero, that is, until the group breaks up

(2) As the externally imposed activity (E) is decreased, then the equilibrium levels of activity (A) and friendship (F) will be decreased (This corresponds in the diagram of Fig 1 to the line $\psi = 0$ moving toward the left)

(3) As E is decreased below some critical value, E_T, F will go to zero, and for some sufficiently small value of E (equal to or less than E_T depending on the location of the intersection of $\psi (A, F, E_T)$ with the A axis) A will go to zero

(4) The level of E required to bring a group into existence is greater than the minimum value, E_T, required to prevent the group, once formed, from dissolution

These, then, are the deductions which Simon draws from the three postulates with which he began They all seem 'reasonable," and seem to be in accord with general observations Some (such as 2) we might have felt to be obviously true from the postulates without the formal derivation Number 1, and especially number 4, are not so intuitively evident

Just what has been gained by making these deductions? It may be possible upon examining them closely to say, 'I knew that all the time But to do this completely misses the point In the beginning, three propositions were set down, what the mathematical model has done is to spell out the logical implications of these propositions It says to the social scientist "You stated only propositions x, y, and z but in so doing you also implicitly stated propositions u, v, and w To reply that it was already obvious that u, v, and w were true as well as x, y, and z is to disregard completely the value of theory What is important is not whether u, v, and w are obvious, but the fact that they are connected by logical implication to x, y, and z A theory in any science gains its utility from such connections, for it is they which allow a few observations to stand for many Otherwise one must go on in a purely descriptive way, never able to predict phenomena u, v, and w from other phenomena x, y, and z

Much so called social theory is of the sort which states interrelated propositions like those of Homans, yet it is seldom that these propositions are examined for their further implications Probably in many cases they are inconsistent, in others they imply absurdities, and in some, like these of Homans, they imply generalizations which on the whole seem to be in accordance with general experience In any case, if they are to contain any of the values of a theory, their implications must be examined

This is not to say that the deductions from these postulates of Homans tell a great deal about the behavior of groups They do not, simply because

the postulates themselves are very weak. What the deductions do tell is first, that the propositions are *not* inconsistent, and second, they tell just how strong a system of propositions Homans has developed from his study of social groups (*)

The Linear Model. Much of Simon's paper is concerned *not* with the above model, but with a model having assumptions of linearity. To the postulates (1), (2), and (3) above are added the conditions that the equations are linear functions of I, A, F, and E. What is the justification for these added assumptions? Do they derive from a more careful examination of Homans' work, or of the data from which Homans developed his generalizations? No, these assumptions are of an entirely different order than the postulates of the previous section. Those were made on the basis of observation and empirical generalization, these are made for purposes of mathematical simplicity.

Linear differential equations are more easily handled than are non-linear ones, particularly when the non-linear equations are left in general form with few restrictions. At the same time, one can make quite the same deductions about the behavior of the system in the neighborhood of equilibrium from linear equations as from non linear equations. In fact, Appendix 2.2 to this section (on page 42) shows an analytic method for making deductions from non linear equations, this method is based on the fact that the non linear system can be reduced in the neighborhood of the origin (and by transformation of the axes, in the neighborhood of an equilibrium as well) to one with only linear terms.

The introduction of linearity assumptions would ordinarily raise important questions about the *validity* of a model based upon them. For example, consider the equation which in the linear model replaces equation (2.1) of the non linear model

$$I = a_1F - a_2A \tag{2.5}$$

In equation (2.5), activity and friendliness contribute independently to the rate of interaction of the group. But might not friendship "act on" or "intensify" the effect of activity in increasing interaction, and activity intensify the effect of friendship?

Such questions as this would be to the point if deductions were made which depended upon the model's linearity. As mentioned above, the only deductions made are those which concern the behavior of the model near

(*) Further remarks on just how this approach to theory building compares with that of others in the field will be delayed until III

equilibrium, and these do not depend upon the model's linearity (as Appendix 2 2 makes clear) To use linear equations when it is known that their use is not justified may be likened to the use of a Centigrade temperature scale in conjunction with the perfect gas law, $pV = RT$ In general, this will give incorrect results, but if one uses only those deductions which depend upon the existence of linear transformations between temperature scales (that is, deductions which concern only *differences* in temperature) and not upon their having the same zero point, Centigrade scales are all right In the same way, the deductions which Simon makes do not depend upon the linearity assumptions

Thus the linearity assumptions neither add nor subtract anything They provide an alternative means of arriving at deductions, but once having this totality of deductions, Simon restricts himself to those which are valid whether the linearity assumptions hold or not Thus if there had existed no method for arriving at deductions from the non-linear equations, the linear ones would have provided a kind of "crutch" by which to arrive at deductions In certain cases, this might be a valuable trick to use for obtaining deductions for which no method is otherwise available

Because the linear model plays only this role, it will not be examined here It is presented in full in the original paper (Simon [1952]) The remaining comments will focus on the relation of the abstract model to concrete behavior of groups of people, that is, the measurement problem

The Measurement Problem: Relating the Model to the Real World. Although the above discussion is couched in 'meaningful terms like activity, interaction, and friendliness, it has remained in the plane of the abstract, without showing the relation of this abstract system to the real world Neither Simon nor Homans examines this problem — the problem of measurement — in detail Simon indicates that the variables in his system, since they refer to a plurality of individuals, are averages or aggregates The variable *F*, for example, may be defined as the average friendliness between pairs of members of the group But of course this does not define *F* in terms of the real world, it only relates a symbol (*F*) from one abstract system to a symbol (friendliness) from another He does not then relate this verbal symbolic system to the real world by indicating how he would measure friendliness

Homans similarly does not see any need to define precisely the verbal symbols "interaction,' "activity," etc His selection of concepts is in part determined by their relative lack of ambiguity He says, for example, concerning "interaction' (which is the concept from which his definition of

"group" derives): "The charm of interaction for some sociologists is that it can be observed rather unambiguously, that it can in fact be counted" ([1950], p.86). But he does not give any instructions for observing interaction, and different investigators of small groups, Bales, for example, would give quite different instructions to a person who asked him how to observe an interaction from those Chapple would give to the same person.(*)

Such questions may seem like quibbling. Certainly the observations Homans reports and the weak propositions that he draws from them will hardly be disputed by social scientists. But the problem remains: these propositions have been put into a formal model, and implications drawn from them; consequently it is necessary to ask, what is the explicit relation of the model to the real world?

Simon and Homans may reason somewhat as follows: These propositions are quite weak, so weak that *no matter how* one defines the concepts within a common, well-accepted area of meaning for each, the propositions will be true. Then the question is, *are* the relations weak enough so that they will hold within an accepted core of meaning for the concepts? This question is clearly unanswerable without knowing just what the "accepted core of meaning" is. To know this, we would have to question each social scientist about his definition of these concepts, and then try to find the "core of meaning" shared by all definitions of the concept.

Another tack, however, can help to answer the question. That is, from the other direction: Just what do the relations, as stated in the formal model, imply about the "core of meaning"? That is, what invariances between two different methods of measuring interaction (and the other concepts) must hold if the following two statements are to hold: (1) Given that the model agrees with certain empirical data when one method of measurement is used, then the model also agrees with these data when the second method of measurement is used. (2) Or given that the model disagrees with certain empirical data when one method of measurement is used, then the model also disagrees with these data when the second method of measurement is used.

To answer the question will not tell what the "core of meaning" of inter-

(*) See R. F. Bales, *Interaction Process Analysis* [1951a], in which he gives detailed instructions for observation, and Chapple [1949]. However, Homans seems to be concerned with a somewhat different concept of interaction than these investigators. Their concern is with discussion groups who are in physical proximity and who interact by initiating verbal participations. Homans is concerned with groups which are not in any fixed physical relation for a given period of time but are more nearly "natural" groups carrying on various activities together. Yet this very difference points up the necessity for making explicit what is meant by a variable like "interaction."

action, activity, etc is, it will only state that *if* one definition, i e, measurement method, does contain this core of meaning, a certain class of others will also This approach may be useful, for if a number of methods of measurement which are all said to measure interaction do in fact satisfy the invariance relations necessitated by the theory, any one can be substituted for another If not, however, then the "core of meaning" exhibited by all of these methods is not large enough to satisfy the weak relations posited in the theory

In the natural sciences measurements in feet, inches, meters, miles, etc, or the various measures of mass, are all within the allowable set of restrictions for the theory of mechanics They differ from one another by only a scale constant Some may be more convenient than others in a particular case, but all will allow use of theory On the other hand, neither Centigrade nor Fahrenheit temperature is within the allowable set with respect to the perfect gas law ($pV = RT$) However, they are within the allowable set with respect to Charles' Law, a weaker law which says $\Delta V = R\Delta T$ (at constant p) To be within the allowable set for $pV = RT$, a temperature scale must have its zero at absolute zero, which these scales do not, but to be acceptable for $\Delta V = R\Delta T$, a scale need only be some linear transformation of the absolute scale, which means its zero point may be anywhere Centigrade and Fahrenheit temperature scales are linear transformations of the absolute scale, so they are acceptable for this latter law, which allows only a fraction of the implications of the perfect gas law

As these examples suggest, the allowable variation in measurement methods may be determined by reference to the form of the theory itself For Simon's theory, the only restrictions on the relations between the variables are, as we have seen, of the form, $\frac{\partial I}{\partial F} > 0$, which simply says that I is a monotonic increasing function of F (*) It is obvious, then, that any measurement scale for F which is a monotonic increasing transformation of F will serve the purpose For if $\frac{\partial I}{\partial F_i} > 0$ (where F_i is a particular measurement of F), and $\frac{\partial F_i}{\partial F} > 0$, then $\frac{\partial I}{\partial F} > 0$, and conversely

Thus, for the general model, the variables may be measured by any

(*) It is not precisely true that the only restrictions involved the first derivative Restrictions (2 4c) and (2 4d) involved the second derivatives rather than the first. Thus restriction would further restrict the class of allowable scales to those related by a linear transformation But since restrictions (2 4c) and (2 4d) only arose as "saturation" assumptions, and were not part of the empirically-derived proposition they will be neglected here

scales which preserve monotonicity, or rank order, among values of the variables. That is, if a group measures higher in A at time 2 than at time 1 $(A_{t_2} > A_{t_1})$, then any alternative measurement scale for A must similarly place $A_{t_2} > A_{t_1}$ for this group.

The linear model poses a different problem. Just as the perfect gas law, which allowed a wider range of deductions than Charles' law, also placed more restrictions on the temperature scales, the linear model places more restrictions on the measurement scales of activity, friendliness, and interaction than does the general model. It turns out that the criterion they must satisfy is invariance up to a positive scale factor. That is, any deductions which are true for F are true for $F_1 = aF$ (where a is positive) and similarly for the other variables, I, A, and E. This may be verified by substituting aF for F in the equations of the model and examining the resulting deductions.(*) Because of the forms of the equations, a will not affect the deductions, but will be absorbed into the coefficients, which are empirically determined.

Thus if the linear model were developed and all the deductions from it used, this would place serious restrictions upon the class of allowable measures Again the advantages of the general model are apparent, for at the present stage of measurement in social science, almost no measures are invariant up to a scale constant, as required by the linear model.

Aggregation. In the above, there is an implicit assumption that measurements were somehow made on groups as such. But an examination of the methods various investigators use to measure the interaction, friendliness, etc. of a group would show that measurements are first made on individuals, and these are then summed or averaged to give a group measure.(††) What does the general model require of the measurements of individuals'

(*) The deductions may be found in Simon ([1952], p 207).

(††) Methods of aggregating individual measurements of interaction (and the other concepts) other than summation could of course be used Simon suggests summation of individual variables to form a group measure, Homans does not make any specific statements about measurement, and in general considers only individual friendliness, interaction, and activity. In a well-developed state of the science, an explicit theory of aggregation would exist, but we are far from that stage. The use of non-additive aggregation would in general imply restrictions on the class of transformations different from those implied by summation.

Recently, proposals have been made to measure certain characteristics of small groups as a unit. Perhaps the most interesting is the use of projective pictures to infer the group's characteristics from its statements about the picture (which is usually of an informal group). Thus a group which mentions the friendliness between members of the pictured group is assumed to have a high degree of friendliness Such a technique in its present stage of precision could hardly satisfy the requirements of a theory like the one being presently discussed, but it perhaps holds some promise toward development of group characteristics without measurement of specific individuals For reports of the use of this technique, see William Henry and Harold Guetzkow [1951], and Murray Horwitz and Dorwin Cartwright [1953]

friendliness, activity, and interaction? That is, *in order that the measures for group properties, I, A, F, and E be monotonic transformations of one another, what requirements do the individual measures have to fulfill?*

Whenever operations of addition or subtraction are carried out on a measure in order to form a 'composite' measure (as in summing individual measures to produce a group measure), then *if monotonicity is to hold among the set of composite measures, linear transformations must be possible between the original measures.* This may be confirmed by substituting for the individual variables F_j in an aggregation equation an alternative measure F'_j, which equals $aF_j + b$. The arbitrary constants a and b must be the same for all group members, and there must at all times be the same number of members over which the measure is to be summed. If the number of members varies, invariance up to a scale factor is required.

This restriction is of course unfortunate; it narrows the set of allowable measures for the general model from those which are monotonically related to those related by a linear transformation.

Now the question may be asked: is the generally accepted 'core of meaning' of each of the concepts large enough so that two different measurement methods of, say, interaction will satisfy the invariances required by the theory? That is, can two methods of measuring interaction be found *such that for any sample of groups one chooses, the values* (I_1) *obtained by* using one method are linearly related to the values (I_2) obtained by using a second? It is extremely doubtful that such measures can be found. Each of the standard techniques of observing interaction (which is the most fully developed concept, in terms of measurement methods, of those under consideration) focuses on a particular aspect of the interaction. This gives them systematic deviations from one another for different kinds of groups. Bales' method, for example, concentrates on the *frequency of statements* made in discussion, while others focus on the *amount of time spent in discussion*. Still others are concerned with non-verbal *interaction*. A good example of differences which occur is illustrated by the next small group model to be considered. Work by Bales and by Stephan and Mischler which is examined there involves two quite different measures of interaction in small group discussions. Bales counts each "meaningful expression" of an individual in the group discussion as an act of interaction, while Stephan and Mischler count each uninterrupted statement of an individual, which may include many meaningful expressions. It is not hard to imagine these measures differing very much as the type of discussion varies from short interchanges to long monologues. Homans, of course, would accept neither of these methods as measuring what *he* means by interaction.

Certainly these diverse methods of measuring interaction do not give results that come anywhere near satisfying the required condition of linearity, or even of monotonicity, with respect to one another That is, even such a weak requirement as that the methods *rank* groups in the same order is difficult to satisfy If the deviations between measurements were due to 'errors' or disturbing factors, this would be a different thing In the natural sciences, measurements of a property using different techniques almost always have some deviation from one another, that is not the difficulty The difficulty is that here the divergences between two methods are systematic, produced by the different definitions of interaction

These requirements of invariance between methods of measurement would not be relevant if Homans or Simon had stated explicitly what was meant by interaction (I), friendliness (F), and activity (A), or had given a method of measurement for which the generalizations were meant to hold It is only because they did not do so that we must assume they depended on some commonly accepted core of meaning for the concepts to supply the necessary invariance But from the discussion above, it seems that the core of meaning is hardly large enough to supply this invariance It seems that any real differences in meaning given to 'interaction' by different investigators will produce differences in measurement too great to satisfy the required invariances, even of monotonicity

But is this not being a little too restrictive? Would not most social scientists assent to Homans' propositions? Probably so And the reason is this We would be implicitly *holding constant* the various factors which might make two measurement methods differ Only groups engaged in the same general *kind* of interaction, groups in which expressions of friendliness took similar forms, and in which the type of activity was alike would be compared We would ordinarily, in fact, think of comparing the same group over a period of time, thus automatically holding constant many of the variations which would cause difficulties in measurement

Perhaps the most reasonable approach, then, would follow the above lines, *restricting the range* of groups considered, so that reasonable measurement methods for these variables would order groups in the same way This is generally the way in which theories have developed first being restricted to quite homogenous classes of phenomena, with many things held constant Only as the underlying processes became clearer can the theory be elaborated to cover a more heterogeneous class of phenomena (*)

(*) Simon in a personal communication on problems of aggregation has suggested this strategy of restricting the range of groups to be compared as a way of minimizing the measurement problem.

Finally, then, how does the matter stand? A recapitulation of the general lines of the above argument will help. First of all, Homans (and Simon, in his formalization) selects rather *unambiguous* variables, such as amount of interaction, number of activities, and strength of friendliness. But even with these, the ambiguity in definition and measurement is so great that he can set down only *very weak* propositions about their relation to one another. The hope is that these propositions are weak enough to encompass the ambiguity in the variables. However, a study of the formal properties of these propositions shows that, in general, they impose surprisingly *strong restrictions* upon measurement of the variables. Upon the group variables themselves, the various measures must preserve order, but upon the individual variables, where the measurements usually originate, invariance up to a linear transformation is required. The partial resolution, then, is to *restrict the scope* of the theory to relatively homogenous sets of groups, or to changes over time in a single group which maintains much the same structure and function. Such restrictions on the range of variability within groups means that different measurement methods will order groups in the same way. Then as measurement of the concepts is refined and understanding of the underlying processes increases, the scope of the theory broadens to include a wider range of variability among groups.

It may be parenthetically added, however, that the approach to theory-construction implied here might turn out not to be the most fruitful one. That is, it may be that rather than to set down rather vaguely defined concepts, then develop propositions relating them, and finally refine the concepts and elaborate the propositions, another way is better. It may be better first to worry about the processes involved, gaining as full an understanding of them as possible, then let these processes determine the definitions and measurement of the variables. That is, let the processes come first and they will dictate precise – perhaps even quantitative – measurement of the variables.

One of the purposes of this whole discussion of the relation of the model to the real world is to show that it is just as important to be systematic and careful at this point as it is to be systematic and careful about the abstract system itself. Simon has confined his work primarily to the mathematical system, Homans has confined his generalizations to a verbal abstract system without relating it carefully to the real world. The remarks above have attempted to show that without careful attention to the problem of linking up such a system with observations on groups, a model is not complete.

This difficulty seems to be widespread throughout present day social science. Most verbal theories or generalizations which form the foundations

of social science as it stands today are far from precisely related to observation Many concepts are far fuzzier in their meanings than those dealt with here, and certainly many generalizations based on them are rendered almost meaningless Yet with all the inadequacies of such a foundation, few investigators would deny that many hypotheses and ideas flow from it It seems to function in many cases to stimulate further research or to show what elements in behavior have been important to previous investigators, rather than serving as a set of well developed laws

One of the values of such systematization of these verbal theories or generalizations as Simon has carried out is to allow comparison between parts of this foundation, verbal systems do not lend themselves to such comparison We are led to make comparisons only when the generalizations or sets of generalizations are precisely stated and easily interpreted, and this is difficult to do, given the ambiguities of the English language

After the other models are examined, this approach will be compared with them (in III) to determine the similarities and differences of the various approaches It is only in the context of other approaches that the essential characteristics of this one, together with its limitations and potentialities, can become clear

Appendix 2.1.

Since his development of Homans' hypotheses into a systematic model, Simon, together with Harold Guetzkow [1955], has undertaken a more complicated task along the same lines, a translation of a set of hypotheses first set forth by Leon Festinger [1950] Because it constitutes the same approach toward mathematical models in small groups as the Simon Homans model, the details of this work will not be examined, only a few remarks will be made concerning it However, there are some differences, and these will be pointed out below

Festinger set forth, somewhat more systematically than did Homans, a set of interrelated hypotheses concerning communication and opinion deviance in small group behavior Festinger's hypotheses are greater in number than are Homans' (eleven to three), and they deal with more variables The model which results from relating these hypotheses to one another is somewhat more complex than the Simon-Homans model, but it too can be reduced to two differential equations

There are important differences between the substance of Homans' theory and that of Festinger's Homans' hypotheses concern the relation over time of three attributes of a group or a relationship the amount of

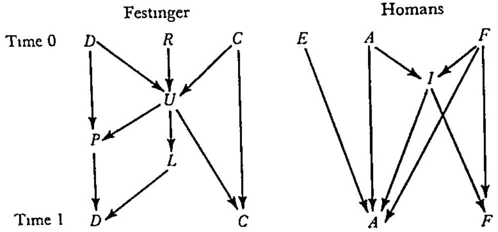
activity engaged in together, the amount of interaction between the members, and the sentiments of friendliness or liking between the members Festinger's hypotheses, on the other hand, concern opinion deviance within the group and variables related to it *pressure to communicate, pressure toward uniformity, receptivity to communications, relevance of the topic to group goals,* and finally, *group cohesion* *Only this last variable, which may be thought of as similar to Homans'* sentiments of friendliness, resembles any *of Homans'* variables Even this similarity, however, is a tenuous one

Homans' hypotheses are concerned with the related growth of friendship and interaction, and the elaboration of activities among the members of a group over a period of time He is concerned with these rather fundamental processes through which people develop relations with and attitudes toward one another Festinger's problem, in contrast, is a somewhat more specialized one, though presumably the processes he is concerned with are applicable to as wide a range of groups as are Homans' *They are processes through which groups come to agree on a subject or to widen their disagreement,* and processes through which they come to feel closer together as a group or to fall apart

The substance of these two theories appears to be different on an even more fundamental level Homans is dealing with two objective conditions surrounding a relationship, conditions which exist "outside of" the subjective state of the members, that is, activity and interaction His hypotheses include only one attribute which characterizes individuals' subjective feelings, that is, *sentiments* of liking or friendliness Festinger, on the other *hand, deals completely with* subjective states of individuals "forces" or "pressures" inside them, and *their opinion of feeling* toward the group His theory, then, is a completely "psychological" one, *in the sense that it* predicts certain psychological states as functions *of other psychological states* In this, it is like other Lewinian theories, which posit *all relationships* "inside" the individual Investigators who use such theories relate them to objective conditions by assuming that some objective change in conditions establishes a change in the subjective states For example, in the experiments on which these hypotheses are based, conditions were manipulated *to produce* what were said to be different amounts of cohesion or of relevance *of the topic* to the group Though it is sometimes easy to manipulate conditions so as to *produce* these changes, this should not obscure the fact that the theory is not *concerned* with the occurrence af such subjective changes, *e g* , changes in relevance *of the* subject or in cohesion, as a function of manipulation Rather, it is concerned with the effect of these subjec-

tive changes, once made, upon other feelings or behaviors of the group members Homans' theory, on the other hand, attempts to relate objective conditions and resulting feelings

The relations postulated in Festinger's theory are considerably more involved than those postulated in Homans', diagrams of the structural relations of the two models show this clearly



D = opinion deviance in the group
P = pressure to communicate
R = relevance of the topic for group goals
U = pressure toward uniformity
L = receptivity of members to communications
C = cohesiveness of group

E = range of external activity imposed on the group
A = range of activity carried out by the group
I = amount of interaction among group members
F = friendliness among group members

The greater complexity of Festinger's theory seems to be due to its subjective nature, it is more nearly based on introspection (supported by experiment) about psychological forces, while Homans' is based more nearly on observation, with less speculation about psychological forces acting inside the individual

Though Simon's treatment of Homans' theory and Simon and Guetzkow's treatment of Festinger's theory do not make evident this sharp formal difference between the two, it is important to recognize that these two approaches to theory building are radically different Summarizing, the difference is between

(a) a theory (like Homans') which relates subjective states to behavior and to external conditions,

(b) a theory (like Festinger's) which relates subjective states to one another and to behavior, and which depends on assumptions outside the theory for relating these subjective states to external conditions

Other differences in the approaches of Homans and Festinger, and the approaches of the two formalizations based on them, are

1 Homans' hypotheses are based on observation of single groups changing over time, Festinger's are based on the comparison of several groups subject to different conditions at one time Thus Festinger's hypotheses are based on inter group comparisons, and his statements that "X increases monotonically with increase in Y mean that a group which is higher in X than another group is also higher in Y As Simon and Guetzkow note, such statements of inter group correlations require added assumptions in order to form the basis of a dynamic system They must be regarded as structural relations, *i e*, X affects Y, rather than covariation, *i e* Y and X are both affected alike by changes in a third variable (*) Furthermore, it must be assumed that the groups are drawn randomly from the same population and that the groups begin from the same initial conditions It is apparent from Festinger's discussion that he intends to make these assumptions, and that he intends the hypotheses to describe stable structural relations which can form the basis of a dynamic system

2 Festinger's hypotheses have been confirmed largely by experiments which manipulate certain conditions, holding others constant, while Homans' are based on field observations, in which the group is subject only to the constraints of the natural setting (†)

3 Festinger's hypotheses include some which refer to a particular (deviant) individual's relation to the rest of the group, while Homans' do not The hypotheses concern the individual s opinion deviance, his being forced out of the group (loss of cohesion between the others and him), the tendency of others to communicate toward him, and other consequences On the

(*) Simon and Guetzkow discuss this difference between structural relations and covariation in some detail

(†) Simon and Guetzkow show that some of Festinger s experiments do not actually test those hypotheses which they are supposed to test Presumably this was due to the lack of an explicit statement of the theory, which would indicate unambiguously the predicted effect of changes in particular variables This is one value of a formalization such as they have carried out — to make explicit the relations between variables so that there is no ambiguity in the predictions of the theory

basis of these, Simon and Guetzkow develop a model which deals with this individual's relation to the group (*). Incidentally, it might be mentioned that such an approach, which essentially considers the group as a "constant" and isolates an individual member for variation, constitutes perhaps the best approximation to a model of small group behavior based on individual level variables as long as the model remains of this general type. Experience in attempting to formulate models with postulates which specify only the sign of the partial derivative in the relationship, like those of Homans and Festinger, has convinced the present author that nothing is to be gained by attempting to begin with individual or pair relation variables and then build up to the aggregate level. Only when the form of the relations is fully specified is it possible to begin with equations characterizing each individual and his relations to the others, and to generate from them deductions about group behavior. Otherwise the postulates are so weak as to leave the situation indeterminate.

These few remarks will suffice to indicate something about the Simon-Guetzkow model based on Festinger's hypotheses. For the reader interested in this kind of model-building, the Simon Guetzkow paper probably provides a better introduction than Simon's paper on the Homans propositions. While the Festinger model is more complicated, the steps of construction are presented more fully, and considerably more space is devoted to discussion of both the substantive and formal aspects of the model.

Appendix 2.2.

The graphically-derived deductions which Simon carried out, and which were followed in the exposition above, may be verified by use of the general analytic theory which has been developed for the study of non linear differential equations (†). Suppose one is given a system of two differential equations,

$$\frac{dF}{dt} = \phi(A,F) \tag{2.6}$$

$$\frac{dA}{dt} = \psi(A,F) \tag{2.7}$$

These can be written,

$$\phi(A,F) = aF - bA - \phi'(A,F), \tag{2.8}$$

$$\psi(A,F) = cF - dA - \psi(A,F), \tag{2.9}$$

(*) This extension of their model is published in Simon [1957].

(†) The presentation of the theory follows N. Minorsky [1947].

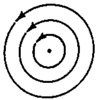
where ϕ' and ψ' are polynomials having neither constant terms nor linear terms in A and F. Then under quite broad assumptions(*) it can be shown that in the neighborhood $A = F = 0$, the investigation of the behavior of the original system can be reduced to that of

$$\frac{dF}{dt} = aF + bA \tag{2.10}$$

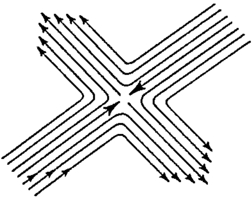
$$\frac{dA}{dt} = cF + dA \tag{2.11}$$

With these equations, one can then examine the *singular points* of the system, which are those points for which $\frac{dF}{dt} = \frac{dA}{dt} = 0$. That is, these are equilibrium points of the system; they may be either stable or unstable, and of the six general types defined and illustrated below:

1. Vortex point: singular point not approached by any trajectory

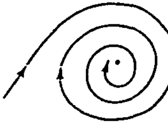


2. Saddle point: singular point approached as shown



(*) The assumption is essentially that ϕ and ψ be analytic (i.e., expressible in a power series in A and F).

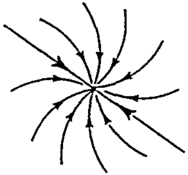
3 and 4 Focal point singular point which trajectories approach as shown



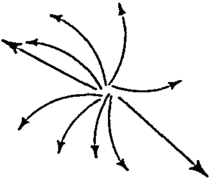
stable

unstable

5 and 6 Nodal point singular point which trajectories approach or leave as shown



stable



unstable

From the figures it is evident that the stable equilibrium (the upper singularity) of Fig 2 3 is a stable nodal point, while the unstable equilibrium (the lower singularity) is a saddle point

If empirical restrictions are placed on a, b, c, and d in equations (2 10) and (2 11), then the theory tells what kinds of singular points the system will contain, if it contains any That is, the theory of these equations states (letting $p = -(a + d)$ and $q = ad - bc$)

If $q < 0$ a saddle point is implied
If $q > 0$, and $p > 0$, and $p^2 < 4q$, a stable focal point is implied
If $q > 0$, and $p > 0$, and $p^2 > 4q$, a stable nodal point is implied
If $q > 0$, and $p < 0$, and $p^2 < 4q$, an unstable focal point is implied
If $q > 0$, and $p < 0$, and $p^2 > 4q$, an unstable nodal point is implied

The empirical restrictions (2 3a), (2 4a), (2 4b), (2 4c), and (2 4d) that were placed upon ϕ and ψ, which determined how Fig 2 3 should be drawn, tell something about the values of a, b, c, and d

$$\frac{\partial\phi}{\partial F} < 0 \text{ implies } a < 0$$

$$\frac{\partial\phi}{\partial A} > 0 \text{ implies } b > 0$$

$$\frac{\partial\psi}{\partial F} > 0 \text{ implies } c > 0$$

$$\frac{\partial\psi}{\partial A} < 0 \text{ implies } d < 0$$

As they stand, these assumptions imply only that $p>0$, which does not tell what type of singular points the system may have This says only that the singular points will be neither unstable focal nor unstable nodal

The "saturation" or concavity assumptions also imply something about the coefficients a, b, c, and d They imply *changing* values of the slopes of the lines $\phi = 0$ and $\psi = 0$, that is, changing values of $\frac{-b}{a}$ and $\frac{-d}{c}$ $\frac{-b}{a}$ will decrease (considering the A axis as reference) as A and F increase, since $\phi = 0$ is concave to the A axis $\frac{-d}{c}$ will increase as A and F increase, since $\psi = 0$ is concave to the F axis As A and F increase, then the final time the two curves cross (*if they cross at all, and if the curves are extended indefinitely*), the slope of the line $\psi = 0$ will be greater than the slope of the line $\phi = 0$ That is,

$$\frac{-d}{c} > \frac{-b}{a}$$

Multiplying both sides by ac (which changes the direction of the inequality, since $ac<0$) gives

$$-ad < -bc, \text{ or } ad - bc > 0$$

Thus $q(=ad-bc)>0$ Also,

$$\begin{aligned} p^2 - 4q &= a^2 - 2ad - d^2 - 4ad - 4bc \\ &= (a-d)^2 - 4bc \\ &> 0, \end{aligned}$$

or

$$p^2 > 4q$$

The restrictions $p>0$, $q>0$, and $p^2>4q$ imply a stable nodal point This corresponds to the upper singularity point in Fig 2.3 Similarly, it can be

shown that if the curves are allowed to extend indefinitely downward, the singularity point (again, assuming the system has at least one such point) at the least values of A and F will have slopes $\frac{-b}{a} < \frac{-d}{c}$ This implies that $q < 0$, meaning that the singularity point at low A and F will be a saddle point, which is unstable This again agrees with Fig 2 3

The values of this analytical approach lie primarily in this by working back from the types of equilibria empirically found in groups, one could begin to set up propositions about group behavior based on these types of equilibria For example, if there were evidence that the activity and friendliness of a certain type of group approached a stable *focal* point rather than a stable nodal point, this would imply that if the variables were related as in equations (2 3) and (2 4), they would have different restrictions from those implied by the propositions which Homans and Simon state These different restrictions would imply different propositions concerning the relation between activity and friendship

It is not necessary to use this analytical method of solution, as we have presented it, it certainly does not aid the intuition as does the graphical method On the other hand, it shows the full generality of the equations, it gives confirmation to the graphical method, and it indicates the empirical restrictions on the equations which correspond to particular kinds of equilibria

3 SOME INTERACTION MODELS IN SMALL DISCUSSION GROUPS

The second set of models to be examined is based primarily on work by R F Bales and his associates [1951a] and by Frederick Stephan and Elliot Mischler [1952] Bales has developed, over the course of several years, a method of recording verbal interactions in small discussion groups His method, which he calls "interaction process analysis, includes coding an 'act" of verbal participation into one of twelve descriptive categories (*) Bales records the participation acts to and from each member of the group in a discussion period, using these categories This gives him the amount

(*) These twelve categories are 1) shows solidarity 2) shows tension release 3) agrees 4) gives suggestion, 5) gives opinion, 6) gives orientation, 7) asks for orientation 8) asks for opinion, 9) asks for suggestion, 10) disagrees, 11) shows tension 12) shows antagonism. See R. F Bales [1950] These categories play no part in the models to be discussed here and will not be elaborated further

and type of participation from each person in the group to each other person Some of Bales' work in mathematical models distinguishes the type of participation as well as the amount, while some of it considers only the amount, regardless of the type of act

Stephan and Mischler have recorded data similar to Bales' They too took small (three to eleven members) discussion groups and recorded the rates of participation to and from each member, but they used somewhat different measurements of participation and interaction than did Bales

Both Bales and Stephan first developed a method of measuring verbal interaction, made their measurements on a number of groups, observed certain regularities, and finally fit a mathematical model to these regularities Bales' groups were of diverse kinds He mentions policy-forming groups, experimental problem solving groups, teams and work groups, family groups, groups formed for counseling, planning, training programs, and others ([1951a], p 1) Stephan's were more homogeneous, all of them tutorial groups of juniors and seniors at Princeton University These investigations are similar to many others which have been carried out in small discussion groups, in which rates of interaction or participation have been measured (see, for example, Chapple [1949], Carter [1951], and Gustafson, [1955]) To our knowledge, however, these efforts by Bales and Stephan have been the only attempts to account for these data through the development of a mathematical model (*)

The measurement of interaction in these numerous groups provided much quantitative data, these data were examined for regularities, and tentative mathematical models were developed to characterize the regularities Following this, an "explanation" of the descriptive model, *i e*, an attempt to provide an underlying rationale, was made

The procedure these investigators have used in developing their mathematical models may be characterized as follows

development of measurement *method* → quantitative observation of interaction in many *groups* → discovery of *empirical* generalizations → development of *descriptive* model to account for data →

→ explanatory rationale for *descriptive model*

This procedure is of course different in several ways from the development of the Simon Homans model, these differences will be explored in III

(*) Gustafson has carried out some statistical tests following the procedure outlined below under the "statistical explanation" section His work is discussed there

It is important to point out here, however, that the model as developed is simply a description of what has been found to be true, not a "theory" in the sense that the Simon Homans model is This difference has important consequences for the use to which the model is put, as will be evident later Since the model is simply descriptive, and is a method of reproducing the data in terms of a few parameters, there is no problem of relating the abstract system to the real world by measurement The original measurement of interaction establishes this relation, for it abstracts from actual behaving groups to give real numbers or vectors The ' mathematical model" is simply an economical (and only approximate) restatement of these numbers The present examination of these models will thus be primarily an examination of their mathematical properties, and of their further implications These models are included in this examination not because they exemplify the usual approach to model building in the social sciences, but because they exemplify a different approach, and one similar in many ways to approaches which have proved fruitful in the natural sciences A common procedure of development in natural science has been the discovery of a regularity or law through some new technique of measurement, generalization of the regularity, resulting in a higher order law, and, finally, explanation of the regularity on the basis of a ' theory "(*)

Bales' Model The first ' model ' of Bales to be examined is really no more than an effort to fit a curve to certain data, with no attempt to give an underlying rationale for it He has not continued work on this model, which he apparently regards as unsuccessful It is included here because it serves as background for more recent work, and also because it was the initial effort in this direction

Bales collected data from groups of three to ten members, and aggregated all the data from groups of the same size as follows the group members were ranked on the basis of their total number of "acts" of interaction during the discussion, and a total number of acts for each rank was determined by summing over all the groups For example, for three man groups, the interactions of all the first ranking members were added, those of all the second ranking members added, and those of all the third ranking members added This gives three numbers the total number of interactions for the first rank, the total number for the second, and the total number for the

(*) James B Conant [1950] presents case studies of discovery in natural science These show remarkably well the relationship between experiment and theory in natural science W P D Wightman [1951] also gives excellent case studies of the history of scientific development

third Besides these overall interaction rates for each rank, the interaction rates to and from each rank by each of the others were determined, to give a matrix of inter-individual interaction rates However, only the overall rates of each rank are considered in this first model These data (in percentages of total participation) for the groups of sizes three to ten are reproduced below

TABLE 3.1

		Group Size							
		3	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	1	44.4%	32.9%	46.1%	43.1%	13.2%	39.8%	49.1%	42.6%
	2	32.6	27.3	22.0	18.8	15.2	16.6	19.0	12.0
	3	23.0	22.7	15.6	14.2	11.9	12.6	7.6	10.0
Rank	**4**		17.1	10.5	11.1	9.9	9.9	5.3	9.1
in	**5**			5.8	7.5	8.6	8.6	4.9	6.0
Group	**6**				5.6	6.3	5.5	4.1	5.3
	7					5.0	4.2	3.8	5.0
	8						2.7	3.7	3.7
	9							2.5	3.3
	10								2.8

Bales plotted these data on graphs and found regularities which he felt might allow the data to fit a simple harmonic curve His data were the above numbers (say $y_{i,n}$ where i = rank and n = group size), and the curve he attempted to fit was

$$y_{i,n} = \frac{S}{i \sum_{j=1}^{n} \frac{1}{j}} \tag{3 1}$$

where S = total number of acts of the group $\left(= \sum_{i=1}^{n} y_{i,n}\right)$

Thus in a three man group, the model would predict that the highest ranking man would get $\dfrac{S}{1(1/1 + 1/2 + 1/3)} = \dfrac{6}{11} S$

Similarly the second man would get $(3/11)S$, the third would get $(2/11)S$ Note that this curve has no constants to be empirically determined Or given only the group size, the equation predicts the proportion of acts by each member (6/11, 3/11, and 2/11 for a three man group)

On simply an *a priori* basis, it would seem that a curve like this could

hardly be expected to fit the data, since it predicts quite independently of any information about the group itself other than its size It would predict the same relative participation rates in a group of children in a nursery as in an industrial conference group A truly universal phenomenon would be necessary to produce invariances that were independent of all group attributes other than size

This equation did not approximate the data closely enough to satisfy Bales Although the criterion of how well the model should approximate the data is not clear,(*) it is evident that the approximations to $y_{i,n}$ given by this function are no closer than those given by many other functions

Stephan's Model. Frederick Stephan took the above data of Bales, which the harmonic curve did not fit well, and found an equation which did fit the data (or that part of the data which Bales had published) First, he fit this equation to Bales' data (Stephan [1952]), then he used the same equation to treat data of his own gathered at Princeton on student tutorial groups there The latter groups ranged in size from four to twelve members, and the data were aggregated just as were Bales', rank by rank Stephan found rather remarkable regularities both in his own data and in those of Bales The first regularity concerned the relative rates of participation within a group He found that except for the leader (defined as the person with highest participation), the rates of interaction of adjacent ranks had approximately a constant ratio throughout the group For example, in a group of four (plus the leader), $p_2/p_1 = p_3/p_2 = p_4/p_3$, where p_i is the proportion of interactions or acts of participation initiated by rank i This regularity was evident both in Bales' data and in his own, though the observational methods differed somewhat The regularity can be expressed as a function of the ratio between interaction rates of adjacent ranks, r_n, and another parameter Stephan writes (†)

$$p_i^* = ar_n^{i-1}, \quad i = 1, \quad , n, \tag{3 2}$$

where n is size of the group excluding the leader, r_n is a fitted ratio for the group of size n, a is the fitted proportion of the total participation by the first-

(*) A statistical test like χ^2 might be used for such a criterion, but a rejection of the model by this (or another) statistical test might easily occur if the amount of data was large even if the model fit it very well The power of the test is greater than one would desire when there is a great deal of data, and less than one would desire when there are very few data Besides this, it is certainly far from clear that the assumptions for a χ^2 test are fulfilled here

(†) The notation used here is slightly different from Stephan's In place of his p_i is p_i^*, with p_i reserved for the actual participation rate.

ranking member (excluding the leader), and p_i^* is the calculated estimate of participation by the i^{th} ranking member. The leader's participation is not accounted for in this way but is obtained by subtracting the p_i^*'s for all the rest of the members from 1.0(*). That is $p_L^* = 1 - \sum_{i=1}^{n} p_i^*$, where p_L^* is the estimate of the leader's participation.

The p_i^*'s obtained from equation (3.2) correspond closely to the actual p_i observed. The observed rates and calculated rates are reproduced below for groups of six (plus leader), from the Princeton tutorial group data.

TABLE 3.2

rank	participations	proportion, p_i	calculated proportion, p_i^*
L	912	45.6	45.8
1	416	20.8	20.1
2	245	12.2	13.2
3	175	8.6	8.8
4	119	5.9	5.8
5	86	4.2	3.8
6	46	2.4	2.5

The fit of the calculated proportions, p_i^*, to the actual ones, p_i, is about the same for the other Princeton groups (Stephan and Mischler [1952], p. 603). For the Bales groups it is not quite as good, but nevertheless seems to be a good fit.

Along with this regularity is a similar one concerning the number of acts of participation or interaction received. Each member of the group is characterized both by the number of interactions he initiated, and those he received. By ranking members according to number of acts initiated, they are also ranked according to number of acts received, that is, the highest initiator is also the highest receiver. Further, the same general equation which relates the initiator's rates also relates the rates of receiving. If q_i^* is the calculated proportions of interactions received, by rank i, then

$$q^* = a\, r_n^{\,i-1},\ i = 1, 2, \ldots, n \qquad (3.3)$$

where n is the size of the group, excluding the leader, r_n is a fitted ratio of the rates of interactions received by adjacent rank, and a is the fitted proportion of the total acts received which are received by the first ranking member. As in the initiation case, the leader's calculated proportion q_L^* is

(*) However, using leaderless three-, four-, and five-man groups Gustafson [1955] finds the generalization to hold for all members.

obtained by subtracting $\sum_{i=1}^{n} q_i^*$ from 1 Again, these calculated proportions are good approximations of the actual proportions received, the deviations between q_i^* and q_i seldom being as much as a percentage point.

Perhaps the most interesting regularity of all those found by Stephan is the relation between r_n for different sized groups He found for his Princeton data that as the group size increased from three to eleven members (not including the leader) the ratio between ranks, r_n, increased in a regular fashion, nearly linearly He gives a linear approximation for these ratios, as reproduced below, which shows little deviation from the actual average ratios, r_n

$$r_n^* = k - b\,(n + 1) \tag{3.4}$$

TABLE 3.3

n	r_n	r_n^*
3	.589	.590
4	611	607
5	623	624
6	661	641
7	667	658
8	668	676
9	694	693
10	710	710
11	727	727

This table indicates the close approximation which the linear relation gives to the ratios computed separately for each size group A similar linear relation, $r_n^{*\prime} = k' - b'(n + 1)$ serves to approximate the ratios, r_n', for interactions received

With slightly less accuracy a_n can be expressed as a function of n to reduce the number of parameters even further Stephan gives the following expression for a_n as a function of n

$$a_n^* = \frac{234}{n - 4} \text{ (if } p_i^* \text{ is to be expressed in percentages)} \tag{3.5}$$

This reduces to four the total number of parameters necessary to reproduce approximately the initiation interaction rates the two noted previously for r_n^*, that is, $k(= .522)$ and $b(= .0172)$, and the two parameters, 234 and 4, for a_n^* A similar reduction can be made for the interactions received, so that altogether eight parameters can approximately reproduce 164 interaction rates

It may be noted before proceeding that if there were no leader in these groups, so that $\sum_{i=1}^{n} p_i^*$ and $\sum_{i=1}^{n} q_i^*$ were equal to 1 0, then the coefficient a would be logically determined by a particular r For if $\sum_{i=1}^{n} ar^{i-1} = 1$, then it follows that $a = \dfrac{1-r}{1-r^n}$ Thus if these groups had been leaderless, so that every person's interaction rate was expressible by the model (as was another group Stephan investigated), a_n would have been directly determined by r_n^*, so that only the two parameters relating r_n^* to n would be necessary This would reduce to two the number of parameters needed for eighty-two interaction rates, or four for 164 rates, truly an economy of expression

The equation or "model" which these data fit is thus a powerful empirical generalization or law about relative rates of participation in discussion groups of the type investigated by Bales and Stephan Logically, it is no different from a generalization of the sort encountered as postulates of the Simon-Homans model "As λ increases, Υ increases " It is of course much more powerful, but it is simply a statement of an empirical law or generalization, not an "explanation or "theory"

Explanation of the Generalization. An "underlying theory " or explanation for this generalization is a next step which Stephan and Mischler begin to carry out They do not state a precise theory but attempt to develop some general ideas about the kinds of properties which might have produced the data fitting the model They propose that each member of a group has a "verbal participation potential," and suggest that 'the data that are yielded by observation of participation reflect average difference in potential' (p 605) They thus attempt to explain the law they have found by differences among individual group members

Stephan and Mischler suggest that when there is (1) a 'free competitive expression' with no imposed restrictions on the members, and (2) no role differentiation of the members, the relative rates of participation will follow their law These two conditions, they suggest, are necessary if the "participation potentials' of members will generate rates of participation fitting their law However, there is no formal theory underlying the descriptive law, though there are these verbal explanations The authors seem to feel that the rate of interaction expressed by their law constitutes an equilibrium state produced by the "competition" of members for the right to speak This

conception is intuitively appealing, and is similar to suggestions made by others (including Bales) on the dynamics of group discussions.

Stephan and Mischler attempted experiments to aid in testing their explanations, but these give rather different findings than those they expected. They measured interaction rates among a group of "high participators" and a group of "low participators," hypothesizing that in each case a higher ratio between adjacent members would be expected because these were homogeneous groups, whose "potentials" were presumably near the same level. This they did not find, but found instead that the high participation group had a *higher* ratio (a more even distribution of interactions) than the ordinary groups, while the low participation group had a *lower* ratio (a more skewed distribution of interactions) than the ordinary groups.

A "Statistical Explanation." We shall now take a somewhat extended aside to suggest another kind of explanation of the regularities which Stephan found.

The "aggregating" or "grouping" of data for a number of groups of the same size, which Stephan and Bales carry out, raises the question of whether this aggregation might itself introduce part of the regularity. First there is the question of whether the constant ratio between ranks might not have occurred if random numbers had been aggregated in a similar way. In other words, suppose the participations were distributed randomly among the n members — might not the regularity have resulted anyway, simply as a consequence of reordering the members by their participation rates and then aggregating? The comparable statistical model is one in which there are n alternatives (n members of the group), each with a probability of $1/n$, and N trials (N acts of participation). Then the alternatives are rearranged with the one having the highest successes first and so on down. More rigorously, the parallel operations in Stephan's observations and in the statistical experiment are: (Since Stephan's model does not account for the leader's participation rate, this will not be considered below.)

Stephan's operations:	**Random number operations:**
1. n individuals in group,	1. n cells each with a probability $1/n$ of being filled on a given trial,
N participations in each session,	N trials at filling the n cells,
m sessions of groups, observing the number of acts of participation, N_i, of each member	m replications of experiment, observing the number of occurrences N_i in each cell for each experiment

2 ranking the members of each group session according to the number of acts of participation of each member, so that $\lambda_1 > \lambda_2 > \lambda_3 > \dots, > \lambda_n$

3 Summing each rank over all sessions, to obtain

$$\sum_{i=1}^{m} \lambda_{1i} = \overline{\lambda}_1, \quad \dots, \sum_{i=1}^{m} \lambda_{n} = \overline{\lambda}_n$$

resulting in $\overline{\lambda}_1 > \overline{\lambda}_2 > \dots, > \overline{\lambda}_n$

4 Dividing $\overline{\lambda}_j$ by $\sum_{j=1}^{n} \overline{\lambda}_n$ to produce p_j, $j = 1, \dots, n$

5 Testing ratios to determine if

$$\frac{p_j}{p_j - 1} = r \quad (j = 2, \dots, r)$$

2 ranking the cells in each replication according to the number of occurrences in each cell so that $\lambda_1 > \lambda_2 > \lambda_3 > \dots, > \lambda_n$

3 Summing each rank over all replications to obtain

$$\sum_{i=1}^{m} \lambda_1 = \overline{\lambda}_1 \quad \dots \quad \sum_{i=1}^{m} \lambda_n = \overline{\lambda}_n$$

resulting in $\overline{\lambda}_1 > \overline{\lambda}_2 > \dots, > \overline{\lambda}_n$

4 Dividing $\overline{\lambda}_j$ by $\sum_{j=1}^{n} \overline{\lambda}_j$ to produce p_j, $j = 1, \dots, n$

5 Testing ratios to determine if

$$\frac{p_j}{p_j - 1} = r \quad (j = 2, \dots, n)$$

The logic of the "statistical explanation" is simply this steps 2 through 5 have precisely the same structure in Stephan's procedure and in the statistical procedure, since this is so then if the final test in step 5 is confirmed in the statistical procedure as it is for Stephan's actual data this would suggest that the structure of step 1 is similar in the two cases That is it would suggest that the data could have occurred "by chance" if we mean by "chance" that all members had equal probabilities of talking at each trial (*)

In accordance with the above experimental procedure experiments were carried out using as the cells the digits 0, 1, 2, … [illegible] ($3 < k < 9$ so that $4 < n < 10$) filling these cells from tables of random numbers [illegible] and $n = 6$ and 15 replications the following ratios were obtained between adjacent ranks

(*) It would be preferable to have the formula for the expected value of [illegible] rather than to carry out random number experiments But these expected values are not easy to compute Arnold Simmel has worked out the equation for the expected value of λ_1 and λ_2 (where $\lambda_1 > \lambda_2$) for the binomial case This is

$$\lambda_1 = \frac{N}{2} \frac{I_{\frac{1}{2}}(\dots - 1, \dots - \dots)}{I_{\frac{1}{2}}(\dots, \dots - \dots + 1)} \text{ and } \lambda_2 = N - \lambda_1 \text{ and } I_p [illegible]$$

is the incomplete beta function [illegible] is as yet unsolved so far as I know

TABLE 3.4

	Experiment		Stephan s groups	
i	$r_{i,i+1}$	$r-\bar{r}$	$r_{i,i+1}$	$r-\bar{r}$
1	796	019 (—)	587	060 (—)
2	846	031 (+)	705	048 (+1
3	833	018 (+)	687	040 (+)
4	873	058 (+)	723	076 (+)
5	729	086 (—)	535	112 (—)
		212		336
	$r =$	815	$\bar{r} = 647$	
	$m =$	15	$m = 15$	

The statistical data seem quite similar to the participation data, differing only in the average value of r The similarity between the purely statistical data and the participation data is quite surprising, as Fig 3 1 shows even more strikingly Even the direction of the deviations of r_i from $\bar{r}$ are similar, negative for the first and last ratios, positive for the rest Similar experiments were carried out for $n = 10$, and they too showed a striking similarity to Stephan's data for 10 member groups, even in the direction of deviation of $r_{i,i+1}$ from the average r Here again, however, the average value of r was higher for the statistical data (782 compared to 692 for Stephan's groups) The average deviat on of $r_{i,i+1}$ from $\bar{r}$ was slightly greater for the statistical experiment than for Stephan's data (082 to 062) (*)

Just where does this leave the problem? In some respects, the statistical investigation seems to have shown a remarkable similarity to the empirical data, in other respects, there is no similarity In summary, there is similarity on these points

1) The ratios, $r_{i,i+1}$ for the statistical data are about as constant as those for the empirical participation data

2) The deviations from the average ratio seem to have a similar pattern to the empirical data negative deviations for the last one or two ratios

3) In some cases, the entire pattern of deviations of the statistical ratios (i e, the shape of the graph) follows closely that of the empirical data

There is difference on this point

1) The statistically-derived average ratios are in all cases higher than the participation data $\bar{r}$'s (This means that the participation rates for members in a group have a more skewed distribution than do the statistically-derived numbers)

(*) The statistical data reported here are not selected from a number of experiments The data presented are all that were obtained The similarities found between these and the participation data are therefore not contrived

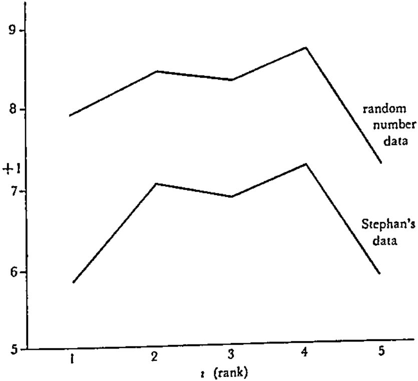


Fig 3 1

And this point has not been examined as yet

1) How does the ratio change with change in n? In the participation data, the ratios increase linearly with group size But what about the ratios from the statistical data? Do they too increase linearly with group size?

Finally, one fact has been purposely not mentioned until now The number of trials, 25, in each replication of the random number experiment, is *not* the same as the number of participation acts in each group session Stephan does not present the N s for separate sessions, but the average, for all sizes of groups, was something near 70 (excluding the leader's participation) Probably most sessions were near this average, for all were 50-minute sessions, and each session was observed for the same part of this 50 minutes

This difference between $N = 70$ and $N' = 25$ is great, perhaps it is the cause of the difference in average ratio between the group data and the random number data That is, if replications of 70 trials, *i e*, random numbers, rather than 25 trials had been taken, perhaps the ratios would have reduced to more nearly that of the participation data But a little reflection will indicate that this increase in number of trials would do exactly the op-

posite, it would *increase* the ratios by making the distribution less skewed. For very large N, the ratios would approach one, with each cell having very nearly $1/n$ of the total. The fact that the distribution of 70 acts of participation is more skewed than is the distribution of 25 random trials must mean that these acts are *not* randomly distributed among the group members, but have some degree of interdependence.

Yet the similarities between the participation data and the randomly-derived data suggest that the participation process is "something like" a set of random trials, except that some association exists between participation acts. It might be, for example, that participation differs from the random trials only in the fact that succeeding acts are partially dependent upon preceding acts. It may be that some such stochastic process is responsible for the distributions of p_i, which are similar to chance distributions, but with some degree of association between the acts of participation.

In order to find out more about the stochastic process which appears to be operating, we may ask the question:(*) Since the (approximately) 70 acts of participation do not give results like 70 random trials, just how many trials *are* they like? That is, what number of random trials would give the same average ratio between rates as that of the 70 interdependent acts? The expected number of such trials, N^*, was computed from Stephan's average ratios for groups from three to eleven (excluding the leader).(†)

(*) An example of the kind of stochastic process which seems to be indicated is given by H C Landau [1951 b] Landau's stochastic process assumes that each event (act of participation in the present case) is the outcome of a two-person encounter, with initially each person having an equal probability of success This differs from the statistical model discussed here, in which it was assumed that each event was the outcome of an n person competition with each having equal probability of success, *i e*, a multinomial model, with each cell equally probable Except for this difference, however, Landau's model seems similar to one which might be used to explain these data, for he assumes (p 249) that each outcome increases the probability of the winner winning on the next encounter and decreases the probability of the loser winning It seems intuitively that some such process might be operative in small group interaction each act of participation by a group member increases the probability that the same member will act again within a short period of time, until some equilibrium of probabilities is reached Landau's model will be discussed in some detail in the next section

(†) The formula for such a computation is

$$N^* = \frac{E\left(\chi^2_{n-1}\right)}{2\left(\Sigma p_i \ln p_i + \ln n\right)} = \frac{n-1}{2\left(\Sigma p_i \ln p_i + \ln n\right)}$$

This hypothetical N, labelled N^*, is based on equating the expected variance of p_i for N observations, $E\frac{\left(\chi^2_{n-1}\right)}{N^2}$, to the estimate of the variance using the actual p_i's, that is, $2(\Sigma p_i \ln p_i - \ln n)(1/N)$; and replacing N by N^* This latter formula is the maximum likelihood estimate of this variance (approximately equal to the more commonly used $1/N\,\frac{\Sigma(p_i - 1/n)^2}{1/n}$).

The following values for Λ^* were found

TABLE 3 5

n	Λ^*
3	7 9
4	8 6
5	8 5
6	10 0
7	9 9
8	9 4
9	10 6
10	11.3
11	11 9

These values are rather constant for all of Stephan's groups from three to eleven members, all the values are somewhere around ten independent participations (*) This suggests that for all these sizes there is nearly the same interdependence of participation (though there seems to be a little less interdependence — more independent acts — as the group size increases) This "interdependence' between the 70 participation acts might be thought of in this way When the first person speaks, this increases his probability of speaking next (or more realistically, the time after next, after the leader or another person has spoken) by some amount This would continue, so that the more a person talked, the greater his chance of talking next, up to some equilibrium point How this equilibrium might be determined, however, is not apparent

These results indicate that the variation in acts of participation by members was such that it was *as if* about ten independent participations had been made, rather than about 70 acts, each having some interdependence with the others This is an interesting finding, one which suggests that the regular increase of r_n with n which Stephan found is simply a result of the constant interdependence of the participation acts as n increases

This is about the limit of the 'statistical explanation There may be other rationales, other chance models, which will go further in accounting for the regularities Stephan found Every other avenue investigated however, has proved not to add to the 'explanation " One particularly important negative result should be stated After finding that the data were "as if" $N = 10$ rather than 70, we conducted more random number experiments for $n = 10$, 6, and 4 with $\Lambda = 10$ The ratios are reported below

(*) Bales data for groups of three to nine (leader excluded) (see Table 3 1) were used also to compute "independent participations " His data correspond to about 9 0 independent participations and are somewhat more constant over change in n than are Stephan's However, Stephan's data show more within-group regularity so they have been used here and elsewhere

TABLE 3.6

	$n = 10$	$n = 6$	$n = 4$
i	$r_{i,i+1}$	$r_{i,i+1}$	$r_{i,i+1}$
1	.726	.743	.667
2	.758	.711	.568
3	.773	.593	.440
4	.883	.812	
5	.867	.385	
6	.770		
7	.300		
8	.333		
9	0		
	$\bar{r} = .590$	$\bar{r} = .649$	$\bar{r} = .558$

There are two points of interest here: first, the last few ratios, $r_{i,i+1}$, seem to decrease very markedly for large n; the last three for $n = 10$ are much below others, and the last one for both $n = 6$ and $n = 4$ is much below the others. This tendency, which was apparent for both the statistical data with $N' = 25$ and for the participation data, is much more marked with $N' = 10$. Thus these statistical data are much less near a constant ratio than were the others.

The second point of interest in these tables is that $\bar{r}_n$ seems *not* to decrease with a decrease in n as does that of the participation data. However, Stephan's r_n were calculated by the use of a weighting formula, and by using this same formula, the statistical $\bar{r}$'s more nearly approximate those of the empirical data.(*) These are presented below:

TABLE 3.7

n	Statistical $\bar{r}_n$	Empirical $\bar{r}_n$
10	.706	.710
6	.681	.661
4	.618	.611

These statistical $\bar{r}_n$ are quite close to the empirical ones, as would be expected, since the N's used in obtaining them ($N' = 10$) were approximately equal to N^*, the number of independent acts equivalent to the 70 interdependent ones.

This concludes the "statistical explanation"; it has helped to suggest how the regularities might have occurred, but it does not at all "explain"

(*) The formula for r_n is $\ln r = \dfrac{AE - BD}{AC - B^2}$ where $A = p_i$, $B = ip_i$, $C = i^2 p_i$, $D = p_i \ln p_i$, $E = ip_i \ln p_i$.

these regularities It shows the similarities of the data to randomly generated data, and shows the consistent areas of difference between the random data and participation data It considers the data from the point of view of a group attribute (interdependence of participation acts) rather than in terms of individual attributes, such as the "participation potential" which Stephan proposes It is a direction of work which appears promising for it suggests that the participation rates might be accounted for by a stochastic process of some kind It suggests that the next steps would be studies of the changes in distribution of participation over time Does the distribution tend to get more or less skewed as the discussion goes on, or does it quickly reach an equilibrium? Careful empirical investigations focused on these questions would indicate the nature of the dynamic model which would account for the data Having such data, it might be possible to develop a stochastic model, without it, only static models like the above random number trials can be tested

Wayne Gustafson [1955] has carried out some experiments with 3- and 4-man groups which show two important results on this point First, he finds that as the number of sessions increases, the relative number of participations of persons in adjacent ranks stabilizes For Gustafson's data (using college students) it appears that by the second session, the relative rates are fairly well stabilized Though Gustafson did not measure changes within each session, his sessions varied in length without varying in relative rates, as Table 3 8 below indicates Thus it appears that an equilibrium is quickly reached, and that the equilibrium "carries over" to subsequent sessions

But this could be true without the participants retaining their same rank Gustafson's results show that except in two instances (group 1, 2nd session, group 2, 3rd session), the participants retained their same rank from the second through the fifth sessions Thus there is not only a stability of relative participation in the group as a whole, there is also a stability of individual participation rate These results are reminiscent of those in animal sociology (see Masure and Allee [1934], and work by Rapaport and Landau discussed in the next section), which finds strict and stable dominance hierarchies established among flocks of hens and groups of other animals But more to the point, it indicates that any dynamic model developed to account for these data be one which leads to an equilibrium which then maintains itself

Gustafson attempted to fit the statistical model outlined above, finding as was found above that the statistical model accounted for the constant r

TABLE 3.8

	Session				
Group 1					
Participant	one	two	three	four	five
1	49 7	43 4	39 0	40 1	———
2	31 8	26 7	31 3	33 1	———
3	18 5	29 9	29 6	26 8	———
N	(173)	(371)	(233)	(272)	
Group 2					
Participant					
1	37 0	38 4	42 0	40 3	38 4
2	41 5	32 4	28 6	31 6	31 9
3	21 5	29 2	29 4	28 0	29 7
N	(284)	(185)	(255)	(357)	(185)
Group 3					
Participant					
1	36 9	44 8	44 0	42 7	35 6
2	37 9	21 1	26 7	24 0	30 4
3	12 6	17 1	17 9	22 2	20 4
4	12 6	17 1	11 4	11 1	13 5
	(198)	(375)	(341)	(333)	(362)

within a group, though as expected the value of *r* in the statistical tests was larger than that in the actual group when the number of trials equalled the number of participations

Keller's Model. Another model using some of Bales' participation rate data has been proposed by Joseph Keller in a comment on a paper by Bales (Keller [1951]) (*) The data with which Keller begins are those published by Bales in his previously-mentioned article [1951c] Besides the overall rates of participation from each member and to each member, Bales published the rates to and from each member by each other member, for the aggregated six man groups The data are presented below

TABLE 3.9

		Received							
	Rank	1	2	3	4	5	6	Group	Total
	1		1238	961	545	445	317	5661	9167
	2	1748		443	310	175	102	1211	3989
Initiated	3	1371	415		305	125	69	742	3027
	4	952	310	282		83	49	676	2352
	5	662	224	144	83		28	443	1584
	6	470	126	114	65	44		373	1192
	Total	5203	2313	1944	1308	872	565	9106	21311

(*) Stephan in an independent comment on Bales' article, treated the data much as does Keller, but did not propose an interpretation of the results as Keller does

The data were aggregated as before by first ranking each member of a group according to his total acts initiated This resulted in a matrix for each session similar to that above In forming the composite matrix, corresponding cells in the individual session matrices were summed There was no reranking by participations received, only the first ranking by total participations initiated was used Yet there are a number of regularities in the data, as Table 3 8 indicates Bales stated these regularities in terms of inequalities [1951c], most of them can be summed up by saying that the 6×6 matrix is continually descreasing to the right and down This general characteristic was noted by Keller, who realized that such matrices are sometimes factorable(*) into a product of two vectors, each of which is composed of real numbers decreasing down (and to the right in the transposed vector) That is, if the above columns and rows 1 through 6 were considered as a matrix P, then a vector of "initiation potentials' times a vector of "reception potentials" may equal P

$$I\,R = P, \text{ where}$$

$$I = \begin{bmatrix} I_1 \\ I_2 \\ \\ \\ I_6 \end{bmatrix} \qquad R = \begin{bmatrix} R_1 \\ R_2 \\ \\ \\ R_6 \end{bmatrix}$$

and P is the above empirical matrix (†) Keller carried out this factorization, obtaining

$I =$ 63 0	$R =$ 63 0
27 7	19 5
21 0	15 8
14 5	10 7
8 8	5 7
6 9	3 2

which, when multiplied, give a resulting matrix P' which reproduces P rather well

TABLE 3 10

$P =$	1	2	3	4	5	6
1		1231	993	675	357	199
2	1745		437	298	159	87
3	1390	437		238	127	71
4	913	282	230		83	48
5	556	171	139	95		28
6	436	135	111	75	40	

(*) Factorization of this sort is possible when all columns are proportional or alternatively, when all rows are proportional

(†) There is the usual problem of the missing diagonals Keller substitutes fictional entries in accordance with one of the usual methods developed for factorization procedures

Comparison of P with P shows the fairly close correspondence of the computed with the actual data

Keller's work indicates that if the group members are characterized by two numbers, representing an "initiation potential" and a "reception potential," the empirical data are accounted for rather well Thus Keller has substituted twelve independent parameters for thirty independent pieces of data (the entries in the matrix P) What is probably more important, he has given an intuitive meaning to these twelve parameters, using them to characterize the group members The specific values of these parameters do not characterize any specific individual in the groups from which the data were gathered, but characterize the "average" individual in each rank

Bales has stated [1951b] that he feels Keller's interpretation in terms of the two participation potentials is "in the right direction," and presumably further work has been done towards testing these interpretations A likely next step would be to examine single sessions, to see if it is possible to factor out these 'participation potentials' to characterize specific individuals rather than ranks Only then can the test be made of whether there is a stable 'participation potential' associated with each individual Even if this is possible, the potential may be specific to the particular group session, and thus not a property which can be used to characterize the individual in abstraction from the particular session (*)

Summary. Despite the different investigators involved, the work examined in this section constitutes a rather well defined unit All these investigators have been concerned with a particular topic, the rates of participation or interaction of the members of small discussion groups Not only is their concern in this single area, but their approaches to the area are similar The models developed are primarily descriptive, only Keller's provides parameters which have some interpretation as individual attributes, and his remains primarily a method of accounting for a set of data with a small number of parameters All this work thus represents a particular approach to theory and research and mathematical model building in the area of small discussion groups Such an approach, which was characterized diagrammatically on page 47, emphasizes the collection of quantitative data, the building of descriptive models to describe the regularities in these data, and finally, the interpretation or explanation of these descriptive models by postulating certain underlying properties

(*) Bales has carried out some further work in this general direction though not specifically to identify the existence of "participation potentials" See especially R. F Bales and Edgar F Borgatta [1953]

The attempt at a statistical explanation of Stephan's data was carried out in order to determine just how much of the regularity exhibited by participation data would also be exhibited by chance data Some of the same regularities were exhibited by this chance data, but some were not More important, the correspondence between the chance data and the participation data suggest that a rather simple stochastic model might account for the data Similarly, the explanations offered by Stephan and by Keller, and concurred in by Bales, suggest work in the direction of locating attributes of individuals which would account for their relative participation

The comparison of this approach to the others included in this series will be carried out in III Certain similarities and differences between this and Simon's work which have become evident in exposition will be treated there For the present, another approach must be examined

An appendix below elaborates the 'statistical explanation of Stephan's regularities presented above

Appendix 3 1

The statistical explanation of Stephan's regularity may also be considered from a somewhat different point of view resulting in a rather surprising regularity

The conceptual model upon which the measure of uncertainty or entropy in information theory is based is a model of selection from among n classes, or location within n classes, precisely the same model upon which the multinomial distribution is based It was a multinomial model which served as the basis for the 'statistical explanation' of the participation data of Bales and Stephan (see page 54), so it would seem intuitively that a measure of uncertainty or information applied to this data might in some fashion add to the explanation developed on the basis of a multinomial distribution

The measure H of 'information or 'uncertainty concerning the location of an element of a system partitioned into n classes is

$$H = -\sum_{i=1}^{n} p_i \ln p \tag{3 6}$$

where p_i = the proportion N_i/N, of elements in class i This measure takes on its maximum value when each of the n classes has exactly N/n elements in it, so that $p_i = p_j = 1/n$ In such a case, the uncertainty as to where any given element is located is greatest It takes on its minimum value, which is

zero uncertainty or complete certainty, when one class has all the elements, and all other $n - 1$ classes are empty When all elements are in one class, then there is no uncertainty about the location of a given element

In a multinomial distribution with n equally likely alternatives, the situation of maximum uncertainty or disorder in the sample is that in which there are a great many trials, so that $N_1/N = N_2/N, \quad , = N_n/N$ In this case,

$$H = -\sum \frac{N_i}{N} \ln \frac{N_i}{N}$$
$$= -\sum \frac{1}{n} \ln \frac{1}{n}$$
$$= \ln n$$

It is intuitively evident that this situation, in which each class has the same number of elements, is the one which gives least information about where any particular element is It is the "most disordered," or "most uncertain," or "has the greatest entropy" For small numbers of trials from a multinomial distribution, however, p_i $(=N_i/N)$ will seldom equal p_j (still assuming the classes are equally likely), because of chance variation In fact, for every finite N, and given n equally likely classes, there will be a particular value of H, less than its maximum, which is the expected value of uncertainty for the sample Whenever there are differences among the p_i for any reason whatsoever, there will be a value of H corresponding to that configuration, a value less than the maximum value of H

Consider now the group discussions in which each person may be characterized by p_i, the proportion of all the acts of participation which he initiates (As in the previous work, the leader's participation will be neglected, and the p_i's of the other members adjusted to total to 1 0) This system can then be characterized by its uncertainty or disorder on the basis of these p_i's, just as can a sample from a multinomial distribution This uncertainty has a precise meaning, namely this given that we know the rank order of participation among group members, H measures our uncertainty concerning which member initiated a particular act of participation As intimated above, if one member talked all the time, there would be no uncertainty, if everyone talked equally, there would be maximum uncertainty The decrease in uncertainty from the maximum due to members' differences in participation is simply

$$\Delta H = H_{\max} - H_{\text{actual}} \tag{3 7}$$

$$= -n(1/n \ln 1/n) + \sum_{i=1}^{n} p_i \ln p_i \tag{3 8}$$

$$= \ln n + \sum_{i=1}^{n} p_i \ln p_i \tag{3 8a}$$

Seeking now only to explain why the difference among the f_i varies as it does from one size group to another (or to put it alternatively, why the ratio r varies as it does with size of group) we can raise this question Does this range of f_i vary in such a way that ΔH, the difference between maximum and actual uncertainty, remains the same? Table 3 11 below presents the values of $H_{\max}$, H_{actual} and ΔH for Stephan's groups, to test this possibility

TABLE 3 11

n	$H_{\max}$	H_{actual}	ΔH ($H_{\max} - H_{\text{actual}}$)
3	1 100	1 012	088
4	1 387	1.249	138
5	1 610	1 413	197
6	1 789	1 572	217
7	1 945	1 675	270
8	2 080	1 742	338
9	2 197	1 849	348
10	2 302	1 933	369
11	2 398	2 005	393

As it turns out, ΔH increases consistently as n increases from 3 to 11 Both $H_{\max}$ and H_{actual} increase with increase in n but H_{actual} increases less rapidly than does $H_{\max}$ so that ΔH increases throughout the whole range The question preceding the table is thus answered negatively, the distribution of participation in a group does not vary with size in such a way that ΔH is constant

If, however, the question is modified to ask about the reduction in uncertainty per group member, that is, $\Delta H/n$, this value turns out to be quite constant, with no systematic deviations

TABLE 3 12

n	$\Delta H/n$
3	029
4	035
5	039
6	036
7	039
8	042
9	039
10	037
11	036

What this means is that the reduction in uncertainty from maximum is the same per group member for each size group This then is another explanation of the variation in the distribution of participation with change in size of group r varies in such a way that the ratio $\dfrac{\Delta H}{\text{group member}}$ will be the same for each size of group

The previous 'explanation seems at first to be of a different order r varies as if there were about 10 independent trials in n multinomial classes, for all n from 3 to 11 But as it turns out, these are simply different ways of saying nearly the same thing Here are the equations for each

$$\Delta H/n = \frac{\ln n + \sum_{i=1}^{n} p_i \ln p_i}{n} \tag{3 9}$$

$$\mathcal{N}^* = \frac{n - 1}{2(\ln n + \sum_{i=1}^{n} p_i \ln p_i)} \tag{3 10}$$

Substituting (3 9) in (3 10), we get

$$\lambda^* = \frac{n}{\Delta H} \frac{n - 1}{2n} \tag{3 11}$$

These two measures, λ and $\Delta H/n$, thus are the reciprocal of one another except for a factor $(n - 1)/2n$ This factor is almost a constant as n varies These measures thus turn out to be nearly logically equivalent This should not be surprising, for it was suggested above that for a given n and $\mathcal{N}$ there is an H which corresponds to the expected value of the multinomial distribution, this is the expected value of uncertainty The values of H_{actual} in Table 3 11 then are evidently approximately the expected values of H corresponding to the λ found in Table 3 5, that is, about 10 Similarly, those values of $\Delta H/n$ in Table 3 12 evidently approximate the expected value of $\Delta H/n$ for a multinomial distribution with equally likely alternatives and $\lambda = 10$, with this expected value evidently being constant for different n over the range of r observed here (*)

(*) It is easy to check whether this constancy of $\Delta H/n$ for a given λ would hold for very large n and it turns out not to Suppose $n = 1000$ then with 10 independent trials the expected vector of ordered cell frequencies would be very near (1 1 1 1 1 1 1 1 1 1 0 0 0 with no more than one observation in each cell To illustrate

$$H_{actual} = -\ln 1/10 = 2\,3$$
$$H_{max} = -\ln 1/1000 = 6\,9$$
$$\Delta H/n = (6\,9 - 2\,3)/1000 = .0046$$

This value is much smaller than those obtained in Table 3 12 for n from 3 to 11 It would become still smaller as n increased However for n smaller than 1000 around 100 and $\lambda = 10$ $\Delta H/n$ seems still to be near that of the values found in the data

It turns out, then, that these groups can be characterized according to the "disorder" of the system or "uncertainty" as to who initiated a given participation More important, it appears that the difference between this uncertainty and the maximum uncertainty possible for a given n is, if divided by the number of group members, approximately constant over different size groups (*) This, of course, does not constitute more than a partial explanation of the regularities found in the data An explanation gives a satisfying 'reason' for a regularity, and to say that the regularity is such that it makes $\Delta H/n$ constant is not to give a satisfying reason But it does give an empirical law, so that if we had data on participation rates of 3-membered groups and wanted to predict the relative participation rates of 11-membered groups which were under similar conditions, we could do so Not only this, but the result that $\Delta H/n$ is constant may be a step toward a satisfying explanation, for it lets one think of the regularity in a new way The results can be thought of in terms of the amount of uncertainty about who is going to speak next introduced by the introduction of a new group member The result says that each new member (from $n = 3$ to $n = 11$) makes the uncertainty such that the difference between the total that he *could* add (the increase in H_{max} as n increases) and the total that he *does* add (the increase in H_{actual} as n increases) is constant

4 RELATIONAL MODELS

Introduction. One area of behavior in groups which has proved most attractive to mathematically inclined social scientists is that which is concerned with certain specified *relations* between people To be sure, much of sociology and social psychology deals with relations between people, but the work to be examined here differs from most other in its explicit and undivided attention to pairwise relations It differs also in that the relations it deals with are restricted to very simple ones Some examples of the types of relations dealt with will indicate why this is so a) the communication relation an individual does or does not communicate with another, b) a sociometric relation an individual does or does not want to sit next to another in school, c) a dominance relation an individual dominates another or is dominated by him In all these examples, the simplicity of the relation lies in its two-valued nature As measured, it is an all-or-none matter, although in actual fact there may be degrees of variation in the relation, such as the amount of

(*) The same constancy of $\Delta H/n$ was found for Bales groups (again excluding the leader) For most n, these were even more constant than those for Stephan's data being about 040 for most sizes

communication or the degree of friendship. By restricting the measurements to two values rather than considering the whole range of variation which may exist, the social psychologist simplifies both the problem of measurement and the problem of mathematical representation.(*) The complexity arises when a number of persons are considered, rather than a single pair. It is this multiplicity of quite simple relations which makes for socially and mathematically interesting problems.

The early attempts to deal with these two-valued relations symbolically used graphs like the one in Fig. 4.1 below:

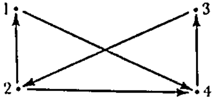


Fig 4 1

Such graphs have been used to represent numerous kinds of relations. In most social applications, the points represent persons, while the arrows represent some feeling of behavior of individual *i* toward individual *j*, or some condition existing between them. The arrow in this diagram from point 2 to point 4 might mean that individual 2 has chosen individual 4 as his friend on a test, that he has been observed speaking to individual 4, that he can defeat individual 4 in a fight, that he has indicated his dislike of individual 4. Ordinarily, all arrows appearing in a single diagram represent the same type of relation, unless there is an indication to the contrary. That is, a particular diagram will represent one single type of relationship within

(*) Because the problem of measurement is quite simple in these models (the relation usually being defined in terms of the measurement), no detailed consideration of the measurement problem will be carried out as was done in the case of the first model examined (the Simon-Homans model) It should be noted here that some work in this general area has dealt with more than two-valued relations For example, there has been sociometric work which deals with liking, indifference, and disliking (in practice, subjects are asked who they would like to sit next to or be with in some other situation, and who they would not like to be with) Also, some sociometric work has differentiated between first, second, and third choices, other work has examined choices by several different criteria at the same time The most fully developed attempt in this direction is the one by Renato Tagiuri, "Relational Analysis An extension of sociometric method with emphasis on social perception" [1952] This examination will not deal with multiple-valued relations such as these, for there has been little success in developing mathematical models using more than two-valued relations A few attempts have been made using matrices with elements 1, 0, —1 to represent three valued relations in a group, but these have been on the whole unsuccessful because such matrices are less manipulable than are matrices with elements 0, 1

the group, representing say, the "friendship structure" or "*communication* structure" of the group. In sum, the diagrams are simply a method of representing certain simple relations between individuals.

These diagrams, or graphs, of the structure of relations within a group are only one way of representing these structures symbolically. To be sure, they are widely used by social psychologists, and mathematicians sometimes find it easier to deduce theorems by using them (in the branch of mathematics called graph theory) than by other representations. Nevertheless, they are often clumsy, as anyone who has worked with sociometric diagrams well knows, and often the most they can do is to give a pictorial representation of the group structure, completely failing to allow analysis of the structure. It is also true that the pictorial representation may be misleading, since the location of points representing individuals is arbitrary. A good example is:

A careful examination is necessary to realize that these three diagrams represent the same structure.

A symbolic representation for these relations which has proved much more useful in the development of mathematical models is a matrix representation which uses 0 and 1 as cell entries. Ordinarily, 1 entered in the cell of a matrix indicates the existence of a relation, while 0 represents its absence. The relational structure represented graphically in Fig. 4.1 could be represented by the following matrix:

$$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

In this matrix, the five entries of "1" take the place of the five arrows in Fig. 4.1, while the "0" entries are in place of missing arrows which could have been drawn. The matrix representation of relations is used by a number of those whose work is to be examined below; it represents their starting point in making mathematical deductions.(*)

(*) Other symbolic representations have been used besides the graphs and the 0, 1 matrices, but these have generally proved less useful. Skew-symmetric matrices with 1, —1 entries have been used for dominance structures, for example. And at least one author has used symbolic logic as a formalization for sociometric relations. Åke Bjerstedt [1956]

Not only can the matrix representation be useful for mathematical manipulation; it can aid in making distinctions between different types of relations. Substantively, the major types of relations considered in the literature are sociometric relations, communication relations, and dominance relations. In the first two types of relations, sociometric and communication relations, the cell entry a_{ij} (that is, the relation of i to j) implies nothing about the entry a_{ji}. Thus communication relations and sociometric relations are mathematically alike, and may be treated by the same models. Dominance relations, however, are different. Dominance of individual i over individual j (which means that $a_{ij} = 1$) implies that individual j does not dominate individual i, that is $a_{ji} = 0$. Thus a matrix representing a structure of dominance relations must have all $a_{ij} \neq a_{ji}$. The elements above the main diagonal give all the information about the dominance structure, for those below the main diagonal are simply the complement of their counterparts above it. Of the two matrices below, the left-hand one could represent a dominance structure, while the right-hand one could not:(*)

$$\begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \qquad \begin{bmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Because of these differences in types of relation, models developed to deal with dominance relations will be examined separately from those developed to deal with the communication and sociometric choice. Most of the work to be examined here is static, in the sense that the models represent no changing system, but simply a state of the system.(†) However, some of the work, to be examined after the simpler static formulations, treats change in the sociometric structure or the dominance structure of the group through time, as a function of contacts between people which can change the relations. Altogether, the following approaches will be considered:

(1) Those which make assumptions about the genesis of the group structure, and prove certain conclusions from these assumptions. These models may be classified into two groups:

(*) In mathematical terminology, the dominance relation is anti-symmetric, while the communication and sociometric relations are asymmetric. All the relations are not transitive, and not reflexive.

(†) The various sociometric indices, such as indices of "sociometric status," or indices of "cohesiveness" based on the choices made within the group, will not be included here, for most of them are arbitrarily defined For one which is less arbitrary than most, see Katz [1953].

(a) "Random" systems, in which one of the assumptions is that choice, communication, or dominance is randomly distributed throughout the group

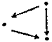
(b) Modifications of this assumption of randomness, to obtain a model which might more easily fit an actual situation

(2) Those which take a structure as it may exist and ask certain questions of it how many 'cliques' are there in the group, how far may communication spread after a given period of time, etc These approaches deal with certain properties of matrices (*e g*, if a matrix is nonsingular, this has certain implications for the group structure)

(3) Those approaches which establish certain postulates as to what will happen through time to change the group structure, and then make deductions about the resulting state of the system at a given later time or at equilibrium

These approaches will be considered in order, dealing first in each case with the sociometric choice or communication relation, and then with the dominance relation

Random Choice Models: Work of Rapoport and Others with Random Nets. Much of the work with random choice has been done under the guise of a quite different kind of substantive problem than that of social behavior Several mathematical biophysicists have worked on problems of "random neural nets," that is, aggregates of neurons which send out one or more "axones" apiece, to other neurons Much of this work is directly applicable to choice or communication relations in groups of people In fact, these authors in some of their writings have suggested and carried out such applications

One of the first problems dealt with by these authors is the probability that a given element in a random net (the element being an individual in our interpretation, neuron in the original work) is a member of a cycle of a particular size An element in a group of four, for example, can be in a 2-cycle $\bullet \longleftrightarrow \bullet$, in a 3 cycle , or in a 4-cycle . As these diagrams indicate, an *n* cycle is a path from the element back to itself in *n* steps by means of the directed relations The work on cycle distributions reported below is that of Rapoport [1948]

If all elements send out only one link, that is, the individuals are restricted to one choice or one communication channel, then the probability

of an element being in a 2-cycle is $1/N-1$. For consider any element i which sends out a link to another element, say j, then j in turn has $N-1$ elements to which it can send its link. Thus the probability of j's link going back to i is $1/N-1$. This is at the same time the probability that i (and j) are elements in a 2-cycle. This result indicates that the probability of being within a 2 cycle in a group in which links are sent out randomly becomes quite small if the group is large, as one might expect.

The general formula for the probability of being a member of a cycle of size k, when each element sends out one link, is

$$C_k = \frac{(N-1)(N-2)(N-3)\cdots(N-k+1)}{(N-1)^k} \tag{4.1}$$

From this formula can be calculated the probability of being a member of a cycle of any size. Since membership in these cycles are disjoint events (an individual can be in at most one cycle, for only one choice is made by each), the probability of being in any cycle is the sum of these probabilities over all k.

$$C = \sum_{k=2}^{N} C_k = \frac{N-1}{(N-1)^2} + \frac{(N-1)(N-2)}{(N-1)^3} + \cdots + \frac{(N-1)!}{(N-1)^N} \tag{4.2}$$

A close(*) approximation to this is

$$C \approx \sqrt{\frac{\pi}{2(N-1)}} \tag{4.3}$$

This, then, is the first problem to be considered: the probability of being in cycles of various sizes in a system in which every element sends out one link randomly. C_k is also the probability that a member will be attached to any other particular member in k steps, and C is the probability that he will be attached to another particular member at all. That is, if two members of the group, i and j, are selected, the probability that there is an attachment from i to j in exactly k steps is C_k, and the probability that there is an attachment at all is C. The derivations for this are exactly the same as for cycles: for there is exactly the same probability of tracing out a path back to a starting point (*i.e.*, a cycle) as of tracing out the same length path to any other previously selected element.

For the moment there will be no question of the use to which such calculations may be put, except to say one thing. It is perhaps obvious that

(*) Rapoport shows that even for N as low as 5 this approximation gives C to within a six per cent error.

there is seldom a group in which choices are given randomly or in which communication or contact is made randomly, so the random model will hardly match the structure of an actual group. The possible value of work like this seems to be its use as a standard against which to measure the structures of actual groups. This possibility will be considered in some detail later.

The above results concern only situations in which each member of the group sends out one link, *i e*, one choice or one communication. When each member sends out more than one link, as would ordinarily be the case in actual experiments or field work, then the problem becomes somewhat more difficult. In such a case, a given element can be a member of more than one cycle, depending upon how many links he and others initiate. This makes the problem of calculating the probability of being within a cycle of length k a more difficult one. However, this problem too has been solved by an approximate method when each member sends out the same number of attachments. Alfonso Shimbel [1951b] shows that for a group of N individuals, each initiating a attachments, the expected number of cycles of k units in length in which an individual will be involved is given approximately by the following equation (*)

$$E(k,a,\Lambda) \approx \frac{(\Lambda-1)^{\Lambda-k-\frac{1}{2}}}{(\Lambda-k-1)^{\Lambda-k-\frac{1}{2}}}\left(\frac{a}{e}\right)^{k} \qquad (4\ 4)$$

This reduces for Λ much larger than k to

$$E(k,a\ N) \approx \frac{a^{k}}{\Lambda-1} \qquad (4\ 5)$$

The value of the latter simplification lies in the fact that often experimenters are concerned with relatively small cycles (two or three units each) in a relatively large group (twenty or thirty members or more). When there is this difference between k and Λ, then the approximation in equation (4 5) is not more than 5 per cent different from the more general result of equation (4 4) (which, it must be remembered, is still an approximation, though a more accurate one).

This result gives an approximation for the expected number of cycles of length k of which an element will be a member. This is not quite a generali

(*) In the original (Shimbel [1951b] p 320), the exponent for $\Lambda-k-1$ is incorrect. Also Shimbel seems to let $\Lambda+1$ equal the size of the group to simplify his notation although he states that the group is of size Λ. As written above in equation (4 4) N is the size of the group and is equal to $n+1$ in his equation (2)

zation of the previously-mentioned "probability of being a member of a cycle of length k,' where only one link was given by each, here the element can be a member of many cycles, depending on the value of a As an approximation, equation (4 4) has been shown to give good results for $a = 2$, $k = 2$, and to reproduce the exact formula for $a = 1$, but how good it is for greater values of k and a is not known It is biased, giving greater values for $E(k,a,N)$ than the exact formula would give

For the generalization from one link to multiple links of the probability that a member will be within a cycle of *any* length, that is, the quantity C, given by equation (4 2), some of these same authors have worked with a notion of 'weak connectivity" This, as they define it, is the probability that an element will be connected to another randomly-selected element, or, alternatively, the probability that the element will be in a cycle of any length This quantity will be labelled C_a Calculation of C_a has been treated in at least three papers, one giving an approximate solution and the other two giving methods of finding an exact solution (Solomonoff and Rapoport [1951], Solomonoff [1952], Landau [1952]) Only the first of these three (Solomonoff and Rapoport) will be considered here, because the exact methods allow a solution for C_a only with great computational difficulty An approximation to the weak connectivity is given by

$$C_a \approxeq \gamma = 1 - e^{-a\gamma}. \tag{4 6}$$

This is a transcendental equation(*) which cannot be solved explicitly for γ, but it can be solved for a

$$a = \frac{-\log(1-\gamma)}{\gamma} \tag{4 7}$$

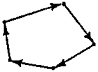
In computing γ from a given a there is not much difficulty if successive approximations are used Note the surprising fact that the value for γ does not depend upon N, the size of the group The approximation is meant to hold only for large N, but one of the authors (Solomonoff [1952]) shows that the approximation is within about 1 per cent of the value given by an exact method even for $N = 4$ and $a = 2$

By examining the value of γ for various values of a the authors show that as a increases from 1, γ increases rapidly, reaching 0 8 at $a = 2$, 95 at $a = 3$, and nearing 1 rapidly for larger values of a

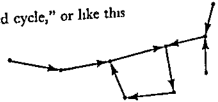
(*) That is an equation which cannot be put into the form of a polynomial set equal to zero

The "strong connectivity" of a group, that is, the probability that there will be a path from a randomly selected element to *every* other element, is mentioned in these papers and given an exact solution in one (Solomonoff [1952]), but the formula is quite complicated, and permits computation only with difficulty.

Besides these properties concerning the likelihood of being in cycles of various lengths, there are certain other properties of such 'random nets" which should be mentioned One has to do with the probability of being in a given position relative to a cycle, though outside it When only one at tachment is made by each member, the possible configurations of the group are much more limited than if there are several attachments If a member is within a k-cycle he can be within no other cycle If he is not a member of a cycle, then he is part of a "tree' which is attached to a cycle That is, all the configurations are either like this



that is, an "isolated cycle," or like this



that is, a cycle with one or more trees' or "tails' feeding in to it, but none leading back out Thus the probability that an element is *not* a member of a cycle of any size (which is equal to $1-C$) is the same as the probability that the element is a member of a tree attached to a cycle

With this as a basis, the probability of being the end element in a tree, and at the same time the n^{th} element from a k cycle can be calculated (*) This probability is equal to the probability of never receiving a link, i e, being an isolate, times the probability of being attached to a k-cycle at a distance of exactly n links The probability of never receiving a link is the probability that each of the other $N-1$ members will send its link to one of the $N-2$ others excluding the element in question This probability is $\left(\frac{N-2}{N-1}\right)^{N-1}$ The probability of being attached to a k cycle exactly n steps away is the probability that the next $n-1$ persons in the 'tree" will not be

(*) This and the other two derivations to follow were made by the author but they are directly analogous to those calculated for a slightly different situation by Shimbel [1948]

a member of any cycle, and the n^{th} person will be a member of a k cycle. This probability is $(1-C)^{n-1}(C_k)$. Putting these two quantities together gives the desired probability, $K(k,n)$ the probability of being the end man in a tree of n links attached to a k-cycle.

$$K(k,n) = \left(\frac{N-2}{N-1}\right)^{\lambda-1} (1-C)^{n-1} C_k \tag{4.8}$$

Similarly, the probability that a man is an isolate, *i.e.*, end man, and after n steps gets to *any* cycle, is

$$K(n) = \left(\frac{N-2}{N-1}\right)^{\lambda-1} (1-C)^{n-1} C \tag{4.9}$$

Equations for several other quantities, without any attempt at showing the derivation, are:

The expected number of isolates

$$E(L) = N\left(\frac{N-2}{N-1}\right)^{\lambda-1}, \tag{4.10}$$

the expected number of trees

$$E(T) = \frac{N^2 C(1-C)}{N-1} \tag{4.11}$$

$$E(T) \approx \tfrac{1}{2}\left(\frac{N}{N-1}\right)^2 \left(\sqrt{2\pi(N-1)}-\pi\right), \tag{4.12}$$

and $E(B)$, the expected number of terminal branches in a tree (or the number of isolates per tree), is the expected number of isolates divided by the expected number of trees

$$E(B) = \approx \frac{2(N-1)^2}{eN\left[\sqrt{(N-1)2\pi}-\pi\right]} \tag{4.13}$$

using the fact that

$$\left(\frac{N-2}{N-1}\right)^{\lambda-1} \longrightarrow \frac{1}{e} \tag{4.14}$$

These probabilities and expected numbers rather fully characterize the group in which each element sends out one link randomly. The probability of a member's being in any possible location in the structure is included among these, and the expected number of each type of configuration (*e.g.*, cycles of various sizes, cycles with trees of various lengths, etc.) is included

or else can be quickly calculated from these quantities (*) The probability of being a member of any kind of cycle, of being at any point in a tree attached to a cycle, the probability of being an end man, receiving no attachment, all these are included in the work discussed above

The more complex case, however, where each element initiates more than one attachment, is less fully handled There, only the expected number of cycles of a particular length ($E(k,a,n)$) and the weak connectivity, i e, the probability of being connected to a randomly selected cycle, C_a, have been calculated in the papers mentioned above However, in another paper Leo Katz [1952] has calculated the exact distribution and also an approximation of the number of isolates, given a attachments by each element These distributions may be found in the original paper, the expected number of isolates, the mean of the distribution, is

$$E(L,a) = N\left(\frac{N-a-1}{N-1}\right)^{N-1} \tag{4 15}$$

where a is the number of choices made by each person

In sum, equations for exact or approximate calculation of the following quantities have been given

C_k = the probability in a group in which each member makes one random attachment, that a member is in a cycle with k steps, or is connected to another randomly selected member in k steps Eq (4 1)

C = the probability that a member of a group (as above) is in a cycle of any length or is connected to another randomly selected member in *any* number of steps Eq (4 2, 4 3)

$E(k,a,N)$ = the expected number of cycles of length k of which an element will be a member, if each member initiates a attachments Eq (4 4, 4 5)

$C_a(\approx\gamma)$ = the probability of a member of a group being in a cycle of any length, or being connected to a randomly selected member in any number of steps, if each member initiates a attachments Eq (4 6)

$K(k,n)$ = the probability of being end man in a tree of n links attached to a k cycle, when there is one attachment per member Eq (4 8)

$K(n)$ = the probability of being end man in a tree of n links attached to any cycle in which there is one attachment per member Eq (4 9)

$E(L,a)$ = expected number of isolates in the group when a attachments are made by each member Eq (4 15)

(*) The distributions of each of these variables (C_k C K etc) have not been calculated in any of the papers discussed here nor in any others of which I am aware

$E(T)$ = expected number of trees when there is one attachment per member Eq (4 11, 4 12)
$E(B)$ = expected number of terminal branches per tree Eq (4 13)

These results cover rather well the structure in which one attachment is made per person But for the multiple attachment structure (when $a>1$, or when the number of attachments made per person varies), the surface is hardly scratched Calculation of the expected distribution of cycles and trees, the probability of being connected to a randomly selected member in exactly k steps, the expected number of trees of various lengths, and others have not yet been carried out

But what would be the value of such further work? Just how would it be used in the study of the behavior of social groups? Everyone ' knows that groups of people in communication or choosing one another do not behave randomly, the very basis of sociometric work is the well documented fact that people do not behave randomly in choosing others But to what *degree* do they behave non randomly? This work begins to allow such a question to be answered That is, the random model may serve as a *standard* in either of two ways

1 It may be the basic assumption which can then be modified in one fashion or another to produce a model which does conform more to actual behavior in a group
2 It may serve as a standard by which to compare actual groups, so that groups could be characterized by their degree of divergence from this random model

Some of the possibilities in these two directions will be considered below, first considering models which modify the randomness assumption, and then considering the uses of the random choice models directly for measurement

Rapoport s Random Nets with Distance Bias An example of the way such random models as those discussed above can be used as a basis for more realistic models of behavior is Rapoport s [1951] introduction of a *distance bias*, modifying the randomness assumption The postulates of this model are similar to those of the previous models except for this single change

(1) Each element initiates a certain number, a, of attachments
(2) The probability that element i initiates an attachment to element j is not the same for all j, but is a function of the distance x_j of element j from element i

(3) The elements exist along a single dimension, so that distance from each element is measured along this single dimension (*)

Given these assumptions, Rapoport could examine a number of deductions He selects one, the probability of a directed connection existing between an element and a particular other element x units away This is analogous for the non distance bias case to C_a, the probability of a connection from an element to a randomly selected other element In that case, the case of pure randomness, the probability of connection to each element was the same, in this case, it is a function of x, the distance from the initiating element

It is reasonable to expect that such a modification as this will come nearer to actual behavior than will the original completely random situation That is, in the sociometric choice interpretation, it is more realistic to think of people as having a certain distance bias in their choice rather than making the choices randomly, that they will choose people whom they feel close to, and that the same people one feels close to also feel close to him A modification of this sort is just a beginning, it is somewhat unrealistic to assume that variations in psychological distance can be described by single dimension models But the model may not be too unrealistic for certain situations, to cite a similar case, a single dimensional random walk model has been used with success to approximate Brownian motion in a fluid, which can never be restricted to a single dimension in actual fact

However, there are other questions besides those about the model's realism which have perhaps more cogency (a) the deduction desired is mathematically difficult and tedious though the postulates (1—3 above) are deceptively simple Thus the model may be too complicated for practical use (†) (b) there are a number of parameters on which the solution depends It may be hard to identify these apart from the data to which the model is applied To the extent that they are not identified independently, then the fit of the model to the data is not tested, for those parameter values are used which make the model fit the data This objection is not, however, at all restricted to this model, it is a problem which occurs with most mathema

(*) This assumption, together with the others makes the model something like a one dimensional random walk For a discussion of random walk problems see William Feller ([1950] p 279 362)

(†) Rapoport [1953a 1953b 1954] has developed a somewhat different approach to the problem of distance bias by assuming that the contacts of two people, who are themselves in contact are largely overlapping This model which he applies to diffusion data with some success is simpler than the one discussed above

tical models in their early stages of development. It is only magnified with this one, for the number of parameters is rather great.

Both these difficulties will be evident upon examining the solution below. If $P(x,t)$ is the probability of an element at x receiving at the end of t links a connection from an element at x_o for the first time, then $P(x,t)$ is given by a recursion formula:

$$P(x, t+1) = [1 - \sum_{j=0}^{t} P(x,j)]\ \{1 - \pi[1 - f\,|\,x - x_o\,|\,\Delta x]a\rho P(x, t)\Delta x\} \quad (4.16)$$

where:

> $f|x - x_o|\Delta x$ = the probability that an attachment initiated by an element at x_o will be made to an element at x, that is $|x - x_o|$ units away.
> a = number of attachments made by each element.
> ρ = linear density of elements along X.

After solving this recursion equation up to values of t high enough that $P(x,t)$ begins to vanish, then the probability of a connection from an element at x_o to a particular element at x is given by:

$$\gamma(x) = \sum_{t=0}^{\infty} P(x, t). \quad (4.17)$$

Presumably $P(x,t)$ ordinarily vanishes quickly enough as t increases so that $\gamma(x)$ can be obtained without tedious repetitions of the recursion formula.(*)

This example will not be carried further. As it stands, it illustrates one way in which the previously presented random models can be elaborated to provide models which conform more to actual behavior. As mentioned above, this is only one method of using the random models as a basis or standard. This method is characterized by modifying the assumption of random choice in some fashion to conform more nearly to behavior. Another method of using the random models will be illustrated below.

(*) A paper which indicates that a quite different kind of approach may be used to deal with the problem posed by Rapoport here is a paper by Edgar Reich, "The Game of 'Gossip' Analyzed by the Theory of Information" [1951] Reich considers a problem in which individuals pass information along a line with imperfect communication (a channel with noise in the terminology of communication theory) The question is what is the probability that the symbol sent out at one end will be the symbol received at any given point down the line? This is similar to the problem of the probability of connectivity between two points along a line, and it seems that Reich's approach, with some modifications, could be used to deal with Rapoport's problem.

Randomness as a Basis for Measurement Models. Suppose the situation arises in which each person in a group is asked to choose one other in the group If these people act as people ordinarily do, there will tend to be more mutual choices than would be expected by chance, there will probably be more three-cycles, and there will be fewer very long chains than would be expected by chance Using equation (4 1), and taking as an example a group of size 35, the expected number of persons who are members of cycles of different lengths is found as follows

$$\mathcal{N}C_k = \text{expected number of persons in cycles of length } k$$
$$= \frac{\mathcal{N}(\mathcal{N}-1)\,(\mathcal{N}-2)\,(\mathcal{N}-3) \quad (\mathcal{N}-k+1)}{(\mathcal{N}-1)^k} \qquad (4\ 18)$$

TABLE 4.1

$\mathcal{N} = 35$	k	$\mathcal{N}C_k$
	2	1 03
	3	1 00
	4	0 94
	5	0 86
	6	0 76
	7	0 65

Now suppose in actual fact it was found that in a group of size 35, four sets of 2-cycles (8 persons altogether), two 3 cycles (6 persons), one 4-cycle (4 persons), and no longer cycles occurred That is, the actual numbers corresponding to the expected ones in Table 4 1 are

TABLE 4 2

$\mathcal{N} = 35$	k	$\mathcal{N}_k$	($\mathcal{N}_k$ = number of persons in a cycle of size k)
	2	8	
	3	6	
	4	4	
	5	0	
	6	0	
	7	0	

Given these data, which differ widely from the random calculations, it is possible to ask the question since the choices were obviously distributed more in small cliques than would be expected by chance, could data similar to this have arisen if the 35 man group were broken up into several *within* which choice were random and *between* which there were no choices? That is, could such data have arisen from n discrete groups of size $\mathcal{N}^*$ (where

$nN^* = N = 35$), with each group choosing randomly within itself, but not outside itself? The question may be answered by use of this fact: For a given k and a given N, the quantity C_k/C_{k-1} has a particular value. That is, C_k is a function of N, and may be written $C_k(N)$. Since this is so, it may be that there is an N^* such that $\dfrac{C_k(N^*)}{C_{k-1}(N^*)}$ will approximate the values $\dfrac{N_k}{N_{k-1}}$ in Table 4.2, and such that

$$n\,[N^* C_k(N^*)] = N_k \tag{4.19}$$

for all k (where $n = N/N^*$, or the number of hypothetical randomly-choosing groups). It can easily be shown from equation (4.1) that:

$$\frac{C_k}{C_{k-1}} = \frac{N - k + 1}{N - 1}. \tag{4.20}$$

Using values from Table 4.2, and solving for N^* in equation (4.20), with $k = 3$:

$$\frac{6}{8} = \frac{N^* - 2}{N^* - 1};$$

$$N^* = 5$$

with $k = 4$:

$$\frac{4}{6} = \frac{N^* - 3}{N^* - 1}.$$

$$N^* = 5$$

Thus both these equations indicate that the size of the hypothetical subgroups, within which choice is random, is 5. The number, n, of these hypothetical groups is $35/5 = 7$. A test of the fit of the model may be made by equation (4.19). Table 4.3 below gives in the second column the values which should equal N_k if the model fit perfectly. These values are calculated from the left-hand side of equation (4.19), with $nN^* = 35$ and $N^* = 5$, using equation (4.1) to find $C_k(N^*)$. The actual values of N_k are listed for comparison.

TABLE 4.3

k	nN^*C_k	N_k
2	8.75	8
3	6.6	6
4	3.3	4
5	0.82	0
over 5	0	0

A comparison of the values nN^*C_k with the actual values Λ_k indicates that the model does fit fairly well, and it is therefore reasonable to say that the 35 group members chose *as if* they were seven five man groups within which choice was random (*) The value of such a statement and the calculations on which it is based is that it can act as a measurement of the group's tendency toward cliques, a measure which not only has intuitive appeal but is based on an explicit model whose postulates have behavioral interpretations

This, then, is one example of a method of characterizing the structure of actual groups by using the random models discussed previously It is in effect saying that the groups act like the random model except in one respect, *e g*, in their tendency toward 'cliquishness The measure of that tendency constitutes the measure by which group structures may be compared Of course in order to use such a procedure as this, as in a measurement model, it would be necessary to go much further than we have gone here For example, the variance of C_k (which would probably be large, especially when k is larger than 2 or 3) is not known, and it would be necessary to know this in order to determine just whether the group s structure is significantly different from a random structure It may be that a measurement model based on this procedure would not be practicable because of such difficulties of testing significance of differences, and because actual numbers of persons in five member cycles, say, must be 0, 5, 10, , while the expected values do not conform to these multiples of five Nevertheless, this example indicates the type of approach which one would use in developing a measurement model using the random model as a basis

There is one other example of this approach of which I am aware in published literature and it is very similar in concept to the one presented above A Swedish geneticist, Gunnar Dahlberg, developed a measure of a population "isolate,' that is, the size of a population within which marriage can be considered genetically random [1947] This measure is based on the realization that if marriage were random, first cousin marriages would occur with a given frequency which depends upon the average number of children and the population size Therefore, if the average number of children in a given population can be estimated, and the number of cousin marriages can be determined from marriage records, the size of the hypothe-

(*) The method of estimating N^* and n here is very crude and better methods could be easily devised It should also be fairly simple to devise a statistical test of significance of the fit to a particular n and N^* In order to develop such "pseudo-random" models as the one discussed here, probably the most important steps are to devise efficient methods of estimation of model parameters and tests of significance for the model s fit to the data

tical population isolate can be determined. This is an important concept in population genetics, and the measure has been used empirically by a number of workers.(*)

The random models which were examined earlier may be modified in a number of ways besides those mentioned. That is, if the assumption of randomness of choice is modified, then the modification can be of many kinds. Some modifications which seem to correspond to actual tendencies in behavior are listed below. In effect, these are illustrations of different kinds of interdependence or non-randomness among choices:

1. If, in a group of individuals $A, B, C, \ldots, N$, individual A chooses individual B, then individuals $C, \ldots, N$ are also more likely to choose B than if A had not chosen him. This assumption would tend to focus the choices on fewer people than would be predicted by the purely random model. In actual groups, such overchoosing of some and underchoosing of others has been found (Moreno and Jennings [1945]). (An alternative model would postulate that $C, \ldots, N$ would be *less* likely to choose B after A's choice.)
2. If, in a group of individuals $A, B, C, \ldots, N$, individual A chooses B, then B will more likely choose A than if A had not chosen him. This assumption would increase the likelihood of mutual choices. This also corresponds to what is found in actual groups; more mutual choices are found than would be expected on the basis of chance alone. (Again, an alternative model would postulate that B would be less likely to choose A.)
3. If, in a group $A, B, C, \ldots, N$, individual A chooses B, and $C, \ldots, K$ choose A (either directly or indirectly), then B will be less likely to choose $A, C, \ldots, K$ than he would otherwise. This assumption corresponds to a kind of "hierarchical" situation, in which people tend to choose only "upward."

These three modifications of the assumption of independence between choices indicate the multiplicity of ways the non-independence can arise. Actual behavior in sociometric choice undoubtedly is composed of several tendencies corresponding to several such modifications. A truly realistic

(*) Another example of this "pseudo-random" model approach was evident in Section 3 (the examination of Bales' and Stephan's work), in the statistical explanation, in which the degree of interdependence of participation acts was calculated. There the random model was a multinomial one with equal probabilities of participations for each member. With this random model as a basis, the question was asked, how many participations of this random sort would be equivalent — in the sense of having an equal variance — to the actual data. For other applications of this approach, see Coleman [1953].

model of group structure might have to take a number of tendencies into account. However, it would probably be nearly as valuable to construct for each tendency separately a model which will allow measurement of the strength of this tendency within a particular group. This seems an important and worth-while direction for model building to take in sociometric work.(*)

This completes the analysis of static relational models which are based on assumptions of randomness, except for some models — to be considered later — which serve as the basis for dynamic models of status hierarchies. It should be noted that all the work which has been presented here has considered only one type of relation: the asymmetric, non-reflexive, non-transitive relation, which is perhaps best characterized graphically as an arrow from one point to another.

The remaining work on relational models to be considered is of two types:

(1) that which examines the implications of particular structural configurations. For example, the implications of a particular communication structure concerning the ultimate diffusion of information, or the diffusion after a given number of messages has been sent. In this work, there are no assumptions of randomness; rather, there is only the single question: given this structure of elements and relations, what does it imply about the resulting behavior of the group?

(2) The final work to be examined is, as was mentioned earlier, that which posits changes through time in the group's structure as a function of contacts between individuals.

Implications of Various Relational Structures. The several papers considered in this section all have this one aspect in common they characterize a relational structure as a matrix, and prove certain theorems about such matrices or carry out operations upon them which give added information about the structure. The earliest of the papers of which I am aware is the one by Telson Wei [1948], which contains a few elementary theorems

(*) Further steps in the directions indicated above (including specific models for tendencies 1 and 2 above, as well as others) have been carried out by the author [forthcoming], in a chapter on Structural Sociology Similar models, which measure the non randomness of some behavior relative to the group structure, *e g*, the homogeneity of behavior between persons who choose one another, have proved of great value in a recent empirical study These models allowed the tracing out of patterns of diffusion in the use of a new antibiotic drug among doctors in four small midwestern communities See Coleman, Katz, and Menzel [1957].

about "neural nets" represented in matrix form. Wei's matrices are composed of $a_{ij} = +1, -1$, or 0, representing positive, negative, and no connections, respectively, from i to j. Wei defines a "simple chain" as any structure in which each element (neuron in his interpretation, person in ours) initiates a single link and each one receives one. Note that this definition is not the same as the definition of "chain" used previously, which was restricted to a continuous chain. Wei's "chain" need not be a single continuous chain, but may be composed of discrete cycles. Wei then shows that:

(1) The N individuals comprise a simple chain if and only if

$$AA' = I, \text{ that is, if } \sum_{k=1}^{N} a_{ik}\, a_{jk} = \delta_{ij} \tag{4.21}$$

$$\text{where } \delta_{ij} = 0 \text{ if } i \neq j$$
$$= 1 \text{ if } i = i.$$

(2) If the rows or columns of a matrix A which represents a simple chain can be rearranged to produce submatrices $A_1, A_2, \ldots, A_r$, so that a matrix of submatrices in the diagonals and zeroes elsewhere results (as

$$\begin{bmatrix} A_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdot & \cdot & \cdot & A_r \end{bmatrix}$$

above), then the simple chain is decomposible into a set of disconnected cycles represented by $A_1, A_2, \ldots, A_r$. More formally, if

$$t^{-1}At = \begin{bmatrix} A_1 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & A_2 & \cdot & \cdot & \cdot & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & A_r \end{bmatrix} \text{ where } t \text{ is a}$$

nonsingular matrix of order n, then $A_1, A_2, \ldots, A_r$ are disconnected cycles.

Because these results are almost self-evident when one examines the effect of these operations on the elements of the matrix, and because the case in which each individual initiates an attachment and each receives one is a very special case, this work will not be considered further. These two theorems can be used in certain special cases to learn something about the group structure, but they are not generally applicable. Some further work of other investigators which is more generally applicable to social networks will be considered next.

R Duncan Luce and Albert D Perry [1949] and, independently, Leon Festinger [1949] treat primarily the problem of describing the number of chains and cliques among members of a group The relations between all pairs of members are characterized in matrix form, and through operations on that matrix, it is possible to show how many chains of a given length or cliques of a given size each element is a part of In essence, this is a mathematical technique, using matrix multiplication, to do what sociometrists would do in tracing out the connections which relate one person to another

Luce and Perry's matrix representation, together with their most important theorems, will be discussed here, to show the general approach Their matrices are the same as those discussed on p 71 They characterize a group's structure by a matrix with N (= number of members) rows and N columns An entry '1" at a_{ij} indicates that there is a directed connection from element i to element j An entry '0 indicates that no connection exists The diagonals, a_{ii}, are 0 as well

Given this matrix representation, Luce and Perry investigate what they call 'n chains' from one person to another An n-chain exists from i to j if there are n directed connections, from i to k, k to l, and so on with the n^{th} connection to j n chains are different if the paths are not identical Luce and Perry show that if the matrix is raised to the n^{th} power by matrix multiplication, then the n^{th} power of the matrix allows one to determine the number of n chains of which each element is a member More precisely, they show that A positive integer D occurs in the ij entry of the n^{th} power of the matrix A if and only if there are D different n chains from i to j

With this theorem, multiplication of the matrix A by itself a number of times allows one to characterize the structure and each element according to the chains which exist As was suggested before, this mathematical technique allows one to do what would otherwise be done in diagrammatic form There is little advantage of this over a diagram if the group is small, but if the group is large, the advantage is great, for it might be practically impossible to trace out the chains via a diagram (*)

However, this result may have disadvantages for some purposes It includes as n chains paths which are doubled back upon For example, for the four person group represented by the matrix

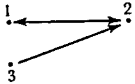
(*) A third alternative to these methods for determining n-chains and n-cycles has been worked out by Coleman and Guttenberg [1956] using IBM techniques This method is applicable to very large groups but becomes cumbersome for chains or cycles beyond five or six members or in cases when the number of links initiated by each person becomes greater than three or four

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$$

the cube of this matrix is

$$A^3 = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

This means that there is exactly one 3 chain each from 1 to 1, 1 to 2, 2 to 1, 2 to 2, 2 to 4, 3 to 1, 3 to 2, 4 to 1, and 4 to 4. In diagrammatic form, the 1 to 2 chain is $1 \longleftrightarrow 2$. The path here is $1 \longrightarrow 2 \longrightarrow 1 \longrightarrow 2$, that is, doubling back upon the mutual linkage. The 3 to 2 chain is



The path here is $3 \longrightarrow 2 \longrightarrow 1 \longrightarrow 2$, again a path which for some purposes it might be desired not to call a three-chain, but simply a one-chain. For other purposes, however, such as the interpretation of the connections as communication paths, these results might be most useful as they stand.

A second result of the Luce-Perry paper concerns "cliques," which the authors define as fully connected sets of three or more elements (i.e., connections in both directions between each pair of elements). A symmetric matrix, S, can be extracted from the original matrix A by replacing each "1" with a "0" whenever $a_{ij} = 1$ and $a_{ji} = 0$. The 1's that are left will represent doubly-connected pairs, i.e., "mutual choices" in sociometric terms. The result concerning cliques is: an element i is contained in a clique if and only if the main diagonal entry of the element i of S^3 is positive. This does not tell which clique the element is within, or what the size of the clique is, but is tells simply that the element is located within a clique.

It should be noted that very little of the information in S^2 and S^3 is used. There are $(N^2 - N)/2$ pieces of information in S^2 and the same number in S^3. The only information used is part of that contained in the N diagonals of S^3, the information as to whether these diagonals are zero or positive. It is very likely, therefore, that further procedures could be developed to use systematically some of the remaining information in S^2 and S^3.

One step in this direction is a result of Luce and Perry's which helps to

identify the number of cliques an individual is within, and the sizes of those cliques This method again depends upon the diagonals of S^3, utilizing their numerical values The values $s_{ii}^{(3)}$ are either zero or positive whole numbers It can be shown that if individual i is a member of one 3 person clique, the diagonal $s_{ii}^{(3)}$ of S^3 will equal 2, if he is a member of a 4 person clique the diagonal will equal 6, if 5 person it will equal 12 In general, if individual i is a member of one m-person clique, then $s_{ii}^{(3)} = (m-1)(m-2)$

If individual i is in two cliques, a three member clique and a four-member clique, then it can be shown that $s_{ii}^{(3)} = (2)(1) - (3)(2) = 8$ That is, $s_{ii}^{(3)} = (3-1)(3-2) + (4-1)(4-2) = 8$ More generally, if no subcliques of three or more members are contained in more than one clique, then

$$s_{ii}^{(3)} = \sum_{j=1}^{r} (m_j - 1)(m_j - 2) \tag{4 22}$$

where m_j is the size of clique j, and there are r cliques in which individual i is a member But this result assumes (for example) that a clique of five of which an individual is a member will not contain three or four of the same persons as another clique of six of which the individual is a member If (1, 5, 7, 8, 9) represents one clique, and (1, 5, 7, 12, 15, 19) another, then the clique (1, 5, 7) of three members is repeated in both In such a case, the value of the duplicated subclique must be subtracted to give the resulting value with which to apply equation (4 22) to determine the number and size of the cliques This subtraction is necessary only for the diagonals of those elements which are contained in both cliques If these are the only two cliques of which individual i is a member, then $s_{ii}^{(3)} = (4)(3) + (5)(4) - (2)(1) = 30$ That is, since a subclique of three members including individual i appears in both of the larger cliques, the $(3-1)(2-1)$ must be subtracted from $\sum_{j=1}^{r} (m_j - 1)(m_j - 2)$ to allow the calculation of clique size and number More generally, the rule is that for each subclique c_j which appears more than once in larger cliques, $(c_j - 1)(c_j - 2)$ must be subtracted from $\sum_{j=1}^{r} (m_j - 1)(m_j - 2)$ each time the subclique appears beyond the first time

One result in a slightly different direction which Luce presents in a later paper [1950] concerns a more generalized notion of cliques and connectivity He defines a set of elements (persons) to be "n connected," if each element i is attached to each other element j in n steps or fewer (considering

attachments to be directed, as before). He then defines an "n-clique" to be one in which n-connectivity holds between each two members. That is, a "2-clique" is one in which each person is at most one step removed from each other person. (Note: this designation of n-cliques should not be confused with the definition of n-person clique. An n-clique may consist of any number of persons greater than two, as long as they are all connected to one another in n steps or fewer.)

The analysis of group structures into n-cliques proceeds similarly to the analysis into ordinary cliques. A symmetric matrix is cubed, and the main diagonal is analyzed. The difference lies in the construction of the symmetric matrix. For analysis into n-cliques, the symmetric matrix, S, is found as follows:

Let G be the matrix of choices. Then construct a a matrix $A(n)$ as follows:

$$A(n) = \| a_{ij} \| = G + G^2 + \ldots + G^n = \| \sum_{k=1}^{n} g_{ij}^k \| . \tag{4.23}$$

Further, let $b_{ij} = 1$ if $a_{ij} > 0$
$b_{ij} = 0$ if $a_{ij} = 0$
$b_{ii} = 0.$

Then extract S from B $(= \| b_{ij} \|)$ so that S is symmetric, letting $s_{ij} = 1$ if $b_{ij} = b_{ji} = 1$, and $s_{ij} = 0$ otherwise. Then the analysis of S into cliques is done as before. The cliques found will be fully-connected cliques of the pseudo-structure, B, but it can be shown that they are the n-cliques of the original structure G.

This is perhaps the most useful result of all this analysis of group structure carried out by Wei, Luce and Perry, Luce, and others.(*) It is particularly useful because it allows a relaxation of the rigid criterion of direct connections in the analysis of cliques. For many purposes, more loosely-bound subgroups such as 2-cliques are desired.

Dynamic Relational Models. Several authors have developed relational models of the sort considered above, except that the structures generated by the models are subject to change. Rather than being a static configuration

(*) Besides the work already referred to, there are two papers by Alphonso Shimbel [1951a, 1952], which review some of this work and suggest several indices of group characteristics, a further paper by Luce [1952] developing the theory of these (0, 1) matrices, and a monograph in graph theory by Harary and Norman [1953] Robert Weiss ([1956] appendix) has developed some of these matrix manipulations into a practical procedure for breaking a larger group into discrete subgroups See also Beum and Brundage [1950].

of choice or communication or dominance, the system may change due to further contacts between the group members In general, these models may be characterized somewhat as follows

(1) An initial configuration of choice or dominance relations is established, either by assuming some initial state or by assuming initial chance contacts
(2) After an interval of time, two members meet, and the outcome of this meeting either changes the previous relation or leaves it the same (If the relation is a dominance relation, the possibilities are either a continuation of the same dominance or its reversal, if the relations are sociometric choices, either member's choice may change independently of the other)
(3) A new structure (or possibly the same) results from this meeting, and after another unit of time there is another pair meeting This continues indefinitely

This is the kind of situation to which these models apply The only added assumption to those models considered previously is that pairs of individuals meet at intervals of time and produce changes in the dominance or choice relationship It will become apparent, however, that the addition of this dynamic element adds a mathematical complexity which makes these models extremely difficult to deal with if there are more than a very few members (that is, 3, 4, 5)

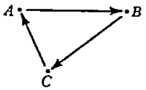
Given the general type of postulates listed above, what are the types of questions characteristically asked by the model builder? They go something like this

What are the possible states after the first meeting, and with what probabilities will each occur?
What is the probability distribution of states after one meeting? After infinitely many meetings?
Is the probability distribution of states after infinitely many meetings independent of the original state?
Under what conditions is there a stable state (i e, an "absorbing" state) from which the system will not move once it reaches this state?

The work to be examined (which is as far as I know, all that exists in the literature conforming to the above characteristics) includes that of three authors Anatol Rapoport [1949a, 1949b 1950] and H B Landau [1951a, 1951b, 1953] on dominance relations and Calvin Leeman [1952]

on sociometric choice Rapoport and Landau have each developed their ideas in three papers, though in neither case does the author devote all his effort in the three papers to dynamic models One of Rapoport's [1949a] and two of Landau's [1951a, 1953] papers concern only the static dominance state, these form a basis for the dynamic models which follow them Some of this work with non dynamic models involving dominance relations will be presented here, both to serve as a basis for the dynamic models, and as a counterpart to the preceding sections which were confined to sociometric choice or communication The dominance relation, with its property of anti symmetry, is quite different from the others, and the random states and various theorems concerning it are different from those

Rapoport's and Landau's Dominance Models. These models have been developed in an attempt to describe the dominance hierarchies among some animal groups A number of investigators had previously studied flocks of chickens and found that a "pecking order" phenomenon occurs, in which a dominance relation or "pecking right exists between every pair of hens (*) Sometimes an unambiguous pecking order was found, with each hen pecking all those below her and being pecked by all those above her Sometimes, however, "cycles" occurred, in which hen A would peck hen B, hen B would peck hen C, and hen C would peck hen A, or diagrammatically



In terms of the types of relations discussed earlier, this means that the dominance relation among these hens was not transitive, that is, it is not true that $A\ p\ B$ and $B\ p\ C$ implies $A\ p\ C$ (where p is the pecking relation) In general, the relation is non transitive, non reflexive, and anti symmetric

One more point is important in noting the results of these experiments The investigators found that this peck right was ordinarily stable over some period of time, but that sometimes, upon meeting, the dominated hen would rise up and reverse her position She would come out of the encounter as the pecker, having entered it as the receiver of pecks In other words, the structure of the flock was subject to change, and not a static order that prevailed

(*) See Rapoport [1949a] and Landau [1951b] for references to these investigations

for all time At all times there was some "turnover," or reversal of position among the hens

Rapoport attempted to build probabilistic models to account for the configurations which existed among these flocks of chickens In doing so, however, he was at the same time building models which could possibly be used to characterize groups of people If instead of a "peck right' some other kind of dominance relation is posited between every two members of a group of people, then these models may be useful for describing the patterns which occur, and explaining the different frequencies of different patterns

Rapoport first attempted to characterize only the static qualities of these structures in his probabilistic models, and later added a mechanism of change His first postulates were

1. Each individual, *e g* , hen or person, meets with each other individual, and the results of these meetings are equiprobable That is, each individual has a probability of 5 of "winning' in each encounter
2. The results of each succeeding meeting are exactly like the first That is, the structure remains the same over time

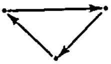
These are very simple postulates, and, as Rapoport suggests, it is hardly reasonable that they will serve as a descriptive model for actual groups, whether of chickens or persons It is more reasonable to expect that the patterns will depend on some attributes of the individuals involved (*) But for a first approximation, and as a standard against which to characterize actual groups, such a model would represent the most egalitarian situation, in which each individual "has an equal chance," with neither his own traits nor a previously acquired position affecting his chances (†)

The questions investigated by Rapoport are simply 1) what are the possible structures in the group? and 2) what is the probability, under the above assumptions, that each will exist?

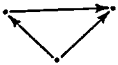
Beginning with a three person group, the possible structures are, in diagrammatic form

(*) Rapoport shows that in order for cycles to exist, the outcome cannot depend on a single characteristic of the individuals involved (unless it does so probabilistically), but must be a function of two attributes of the individuals In fact (although Rapoport does not indicate this), the individual's superiority in an encounter must be a function of one of his own attributes and one of the other individual, if cycles are to occur, and if the relative sizes of abilities determine the outcome

(†) This does not mean that all individuals will end up with exactly the same dominance status, because the chance outcomes can lead to a non-egalitarian structure But at least in large groups, it does mean that a relatively egalitarian structure will be likely



(1)



(2)

That is, in a three man group, it is possible that each individual dominate one and be dominated by one [structure (1)] or that one person dominate two, the second dominate one, and the third dominate none [structure (2)] The probability that each of these structures will occur is easily computed Let the individuals be so labelled that the first encounter is between A and B, and that A is the winner and B the loser The second encounter is between A and C, the third between B and C Then

$$Pr\ [\text{structure } 1] = Pr\ [C \longrightarrow A] \quad Pr\ [B \longrightarrow C] \qquad (4\ 24)$$
$$= 5 \cdot 5 = 25$$

and

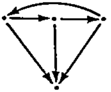
$$Pr\ [\text{structure } 2] = Pr\ [A \longrightarrow C] + Pr\ [C \longrightarrow A] + Pr\ [C \longrightarrow B] \qquad (4\ 25)$$
$$= 5 + 5 \cdot 5 = 75$$

This constitutes the solution (for 3 man groups) of this first problem posed by Rapoport Given the postulates on pages 80 81, Rapoport asked what the possible different structures are, and what is the probability of occurrence of each There are two structures, and the first will occur with a probability of 25, the second with a probability of 75

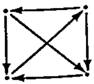
It is evident that for very large groups it would be quite difficult simply to enumerate the possible dominance structures, and much harder to determine the probability of occurence of each structure The number of possible structures is small as long as the size of the group is quite small, but it increases rapidly For example, another author (Davis [1953], [1954]) has shown that when the group is of size eight, there are 6880 distinguishable structures (*) For four individuals, the following patterns are possible

(3 2, 1, 0)

(3 1, 1, 1)



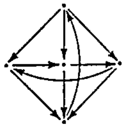
(2 2 2 0)



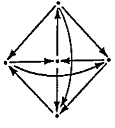
(2 2 1 1)

(*) This work is discussed further on page 106

The numbers below the patterns represent the number of persons dominated by each individual in the group, ordered so that the one who dominates most is first. Such sets of numbers are used by Rapoport and Landau (and by Leeman, in his sociometric model, with a slightly different meaning since he is dealing with sociometric choices) to describe the structure of the group. For groups larger than four, one set of such numbers may characterize different dominance structures, so these do not give all the information about a group's structure. For example, for a group of five members, the set of numbers (3, 2, 2, 2, 1) may characterize either of the following two structures:



(3, 2, 2, 2, 1)



(3, 2, 2, 2, 1)

The importance of distinguishing these two different structures (which will be called "dominance structures" henceforth, while the set of numbers derived from them will be called "score structures") is this: when a mechanism of change is postulated, then the transition probabilities which transform the dominance structure on the left to other structures are different from the transition probabilities which transform the structure on the right to others. In fact, by a single reversal (between the top and the bottom man) the structure on the left can be changed into one having a score structure (4, 2, 2, 2, 0), while no single reversal in the right-hand structure can produce this score structure.

Despite the fact that these sets of numbers do not characterize unambiguously the group's dominance structure, they are useful in characterizing the amount of hierarchization in the group. Clearly, in the earlier example of a four man group, the left hand dominance structure, with score structure (3, 2, 1, 0), is the most hierarchized, and the right hand one (2, 2, 1, 1) is the least, while the other two are intermediate.

Rapoport [1949a] has given a means for computing the probabilities of the various score structures which can arise, and has computed these probabilities for $N = 4$ and $N = 5$. In a later paper Landau has computed the probabilities for $N = 6$. The procedure is quite complicated, however, and

will not be presented here. Landau [1951a] has developed what is probably of more use: a "hierarchy index" based on these score structures.(*) This index, varying from zero to one, shows the degree of hierarchization of a group characterized by dominance relations. The expected value of the index, under the assumption of equiprobable outcomes of meetings, shows the expected degree of hierarchization for groups of any size. The index is:

$$h = \frac{12}{(N-1)N(N+1)} \sum_{i=1}^{n} \left(v_i - \frac{N-1}{2}\right)^2 \tag{4.26}$$

where the v_i are the numbers which make up the score structure. Its expected value for the assumptions on page 80 is·(†)

$$E(h) = \frac{3}{N+1}. \tag{4.27}$$

An examination of the way this expected value varies as N increases makes it evident that under the assumption of equiprobable outcomes, the hierarchization of a group should become very small as N becomes large. For example, if the group is of size 29, then

$$E(h) = \frac{3}{29+1} = .1, \tag{4.28}$$

that is, the expected hierarchization is near zero, far from a complete hierarchy. On the other hand, for quite small groups, structures near the complete hierarchy are more likely. For $N = 4$,

$$E(h) = \frac{3}{4+1} = .6. \tag{4.29}$$

These expected values may be quite useful as standards by which to compare actual groups. Suppose, for example, that among a number of actual groups of different sizes, in which the dominance relation exists, the score structures give hierarchy indices near to the expected values. This

(*) An alternative hierarchy index is presented in Appendix 4 1 This index is based on the disorder of a given structure (which is least when all persons are equal and greatest when there is the greatest possible hierarchy) The index is logically related to Landau's

(†) Landau [1951a] also gives the expected value of h in the general case, when the probabilities of i dominating j in an encounter (p_{ij}) are not equal to 1/2 for all i and j He gives as well an equation for the variance of h

would be evidence that the assumption of equiprobable outcome is valid in these groups On the other hand, values of h which were consistently above the expected values would be evidence that some factor was operating to make certain persons more likely to win in encounters

Landau notes that flocks of about ten hens have generally been found to have score structures near to a complete hierarchy This indicates that the assumption of equiprobable outcomes is far from valid with these hens, and suggests that some attributes of the individual hens account for the outcomes

Landau [1951a] shows some of the characteristics of the dominance structure if the outcomes are assumed to be dependent on a set of uncorrelated attributes or 'abilities' of the group members That is, an individual whose "abilities" are greater than those of another will win in an encounter with that other Landau shows several results, such as the expected values of h under different assumptions about the number of different abilities and the distribution of these abilities in the group The primary result, however, is this if these abilities are uncorrelated (or even if the correlation is positive but not very high) and if the outcomes of encounters are dependent upon them, then the dominance structures which would result are less hierarchized than those actually found among flocks of hens In other words, the empirical results which have been found with flocks of hens cannot be explained by reference to two or more attributes of the individual hens by which they could win encounters Other assumptions which have been introduced to account for such hierarchization will be mentioned shortly when mechanisms of change for these structures are discussed

Until this point, the work discussed has not included a mechanism of change, the structure is regarded as determined once and for all by the initial meetings of each two individuals What happens if upon a second meeting there is a certain probability of reversal? These authors have treated this question by use of several alternative assumptions These will be examined below

Rapoport sets up the process of change as a Markov process,(*) that is, a stochastic process in which the state at time t is dependent on the state at time $t-1$, but on no state previous to time $t-1$ If the probability of being in state i at time t is represented by $S_i(t)$, then the probability distribution over all the possible states can be represented by the vector,

(*) He does not make reference to Markovian theory and proves independently the basic theorems he needs However, the process is a Markovian one and the standard methods used there for obtaining stationary states and other deductions are applicable

$$S(t) = \begin{bmatrix} S_1(t) \\ S_2(t) \\ \cdot \\ \cdot \\ \cdot \\ S_k(t) \end{bmatrix}$$

where there are k possible states. In general, after an encounter, the group may go from dominance structure i at time t to j at time $t+1$, and the probability that i will go to j may be given by s_{ij}. The matrix of these s_{ij}'s or "transition probabilities" is the operator which carries $S(t)$ to $S(t+1)$. That is,

$$S_1(t+1) = S_1(t)\, a_{11} + S_2(t)\, a_{21} + \dots + S_k(t)\, a_{k1} \tag{4.30}$$

$$S_2(t+1) = S_1(t)\, a_{12} + S_2(t)\, a_{22} + \dots + S_k(t)\, a_{k2} \tag{4.30a}$$

$$\vdots \quad \vdots \quad \vdots$$

$$S_k(t+1) = S_1(t)\, a_{1k} + S_2(t)\, a_{2k} + \dots + S_k(t)\, a_{kk} \tag{4.30b}$$

or in the shorthand matrix notation:

$$S(t+1) = A'S(t) \tag{4.31}$$

where

$$A = \begin{bmatrix} a_{11} \dots a_{1k} \\ \dots\dots\dots \\ a_{k1} \dots a_{kk} \end{bmatrix}.$$

Rapoport shows that in such a situation as he has postulated(*) there is a limiting distribution which is independent of the initial distribution. That is, after a length of time, an equilibrium state is reached, in which the probability of each dominance structure remains the same from that time on.

The fact that the distribution does not change over time once it has reached equilibrium may be used to find the equilibrium distribution. That is,

$$S(\infty) = A'S(\infty) \tag{4.32}$$

where $S(\infty)$ is the equilibrium distribution over the possible structures. Using this, it is possible to solve for the components of $S(\infty)$ in terms of the elements of A. For example, in the three-man group it turns out that, by using equation (4.32):

(*) More precisely, the restrictions on A are a) $\sum_{j=1}^{k} a_{ij} = 1$ by definition of a_{ij}'s; and b) if $a_{ij} = 0$, then for some sequence of a_{ik} which carries state i to state j indirectly, the product of this sequence is not zero

$$S_1(\infty) = \frac{a_{21}}{1 - a_{11} + a_{21}} \tag{4 33}$$

$$S_2(\infty) = 1 - S_1(\infty) = \frac{1 - a_{11}}{1 - a_{11} + a_{21}} \tag{4 34}$$

In this case, as in any other, the transition probabilities, a_{ij}, are not the elements with which one begins in developing the model The postulates concern probabilities of encounter and probabilities of one man dominating another, given an encounter, while the transition probabilities are proba bilities of shift from one dominance structure to another Any particular shift may result from a number of different encounters Thus the matrix of transition probabilities is not explicitly part of the postulates, but may in fact be rather difficult to determine Rapoport does not go into this problem, except in the one example he treats In general, he takes the a_{ij}'s as given and deals only with the problem of finding the equilibrium state, that is, $S(\infty)$ However, he treats one example, for a 3 person group, in which he computes the a_{ij}'s from the postulates concerning outcomes of encounters This is presented below

In the previous example of equiprobable outcomes (postulates on page 94) suppose the second assumption is changed to read

2[1] On subsequent encounters after the first between each pair, the probability of each man winning is 5, independent of the previous dominance status of the two

Given this assumption, and given the two dominance structures in a 3 person group

then it is easily shown that $a_{12} = 1/2$, $a_{11} = 1/2$, $a_{22} = 5/6$ and $a_{21} = 1/6$ That is, the probability of going from structure (1) to structure (2) is 1/2, and the probability of going from structure (2) to structure (1) is 1/6 Any encounter in structure (1) will change the structure to (2) if the encounter results in a reversal (which it will with probability 1/2) Thus $a_{12} = 1/2$ Once in structure (2), however, only one of the three possible encounters (each occurring with probability 1/3) can result in a change to structure (1)

if the relation is reversed Thus $a_{21} = 1/3 \quad 1/2 = 1/6$ Using the transition probabilities together with the equations for $S_i(\infty)$ which Rapoport has derived (equation (4 23) and (4 24)), it is possible to find the equilibrium distribution, that is, $S(\infty)$

$$S_1(\infty) = \frac{a_{21}}{1 - a_{11} + a_{21}} = \frac{1/6}{1 - 1/2 + 1/6} = 1/4 \qquad (4\ 35)$$

$$S_2(\infty) = 1 - 1/4 = 3/4 \qquad (4\ 36)$$

$$S\ (\infty) = (1/4,\ 3/4) \qquad (4\ 37)$$

The equilibrium distribution turns out here to be the same as the initial distribution (see page 96) This results from the nature of assumption (2′) above If this assumption had been somewhat different, then the equilibrium distribution would have differed from the initial distribution But what kind of changes in assumption (2′) would produce a change in the equilibrium distribution and what kind of changes would they produce? This is a question asked by Landau in a paper [1951b] in which he investigates the effects of certain "social factors" The social factors are two a "social lag," so that there is a bias against the reversal of an already established dominance relation, and dominance status itself, which acts so that a person who dominates many others will be likely to win in an encounter with someone who dominates few others

The Effect of Social Lag. One reasonable variation to make in the assumption of equiprobable outcomes is this if a dominance relation exists between i and j, then a meeting between the two will tend not to result in a reversal More precisely if $p_{ij}(0)$ is the probability that i will dominate j at time 0, then if it actually comes to be the case that j dominates i at time 0, $p_{ij}(1) = (1 - \varepsilon)p_{ij}(0)$ (where $0 < \varepsilon < 1$) That is, there is a certain bias, ε, against reversal of the pre existing dominance relation

Landau shows an important result for this situation, one which is not *a priori* evident The existence of a bias, ε, and the amount of this bias make no difference in the equilibrium distribution (*) The larger the bias, the longer it will take to reach this equilibrium distribution, but the resulting distribution is independent of whether there is or is not a "social lag" bias, ε

(*) Of course if the lag were great enough, the movement toward equilibrium would be so slow that an apparent equilibrium might exist

This result makes it clear that no such assumption of social lag would change the ultimate hierarchization of the group; the nearly complete hierarchies found among flocks of hens cannot be accounted for in this way. The one factor which such a bias can account for is added stability to the system, with fewer changes in structure than would be found otherwise.

The Effect of Previous Dominance Status. Suppose a somewhat different assumption is made: that the likelihood of reversal depends upon the difference in dominance scores, that is the v_i which go to make up the score structure $(v_1, v_2, \ldots, v_N)$. Landau makes the following assumption:

$$Pr\,[i \text{ dom } j \mid \text{encounter}] = p_{ij} = \tfrac{1}{2}[1 + \omega(v_i - v_j)] \qquad (4.33)$$

$$\left(\text{where } 0 < \omega \leqq \frac{1}{N-1}\right).$$

This assumption is equivalent to saying that the difference in dominance scores $|v_i - v_j|$ augments by some amount (dependent on ω) the probability that the person with the larger dominance score will win in the encounter. This amounts to a linear bias in favor of those with high scores.

As might be expected, such a bias in favor of those who have high scores results in a tendency for hierarchical structures to result. Yet, even if such a bias is at its maximum value $\left(\frac{1}{N-1}\right)$, the expected value of h is far from its maximum except for very small groups. For $N = 3$, and a maximum bias, then $E(h) = 1$; but for $N = 5$, $E(h) = 4/5$, and for large N, $E(h) \approx 6/N$. This means that this assumption of bias in favor of persons with high scores as a linear function of the distance between scores cannot in general account for the almost complete hierarchies found among flocks of hens. Applied to other types of situations, such as continuing round robins in which each person meets each other person over and over again, this assumption might more nearly correspond to the data. In fact, this model could be applied to group discussion situations such as were considered in the previous section, and the linear bias in favor of high scores might be applicable there. In such situations, each participation of individual i directed toward individual j could be constructed as an "encounter" between i and j with i dominating. Such an interpretation might or might not be a reasonable analog to the mathematical assumption. In any event, it is a possible interpretation, and it suggests one possible stochastic process for the group discussion situation.

a situation which certainly seems amenable to treatment by stochastic processes.

In a final attempt to set up assumptions which can lead to a complete hierarchy, Landau suggests the following: any encounter between individuals whose scores (v_i and v_j) differ by more than one will always result in victory for the individual with the higher scores. That is, there can be reversals only in favor of individuals with higher scores except in meetings between persons whose scores are adjacent. For example, in a group with the score structure (4, 3, 3, 2, 2, 1), a meeting between individual 1 (with $v_1 = 4$) and individual 4, 5, or 6 (with $v_1 = 2,2$, or 1) would result in individual 1 winning with certainty. This would lead to $v_1 = 5$ if individual 1 had been dominated before by the person he met, to $v_1 = 4$ if he had not. It could never reduce his score. Only persons with adjacent scores (4 and 3, 3 and 2, or 2 and 1) could exchange scores in the encounter, and it is obvious that this would leave the score structure just as it was. Therefore, the only direction the score structure can change over time is toward a hierarchy. It is evident that over a long period of time a complete hierarchy will result. This does not mean, however, that individuals cannot "work their way up (or down)" in the structure, though the score structure itself remains stable. Since there can be reversals between persons with adjacent scores, individuals may shift in single steps to quite different positions, even after a complete hierarchy is established.

Further Work on Dominance Structures. Besides setting up models which lead to certain dominance structures, as Rapoport and Landau have done, various authors have obtained several additional results which have relevance for these models. In particular, there are three theorems about these structures which may be of aid in further work with these models.

Landau has shown [1953] that the group member with the largest dominance score (*i.e.*, the member who dominates most others) dominates every other person in the group either directly or through an intermediate person. More precisely:

If v_i is the largest score obtained by any member, then for any member j ($\neq i$), either i dominates j or there is a member k such that i dominates k and k dominates j.(*)

This is a rather surprising theorem, since it holds not only in the complete hierarchy (where, in fact, the member with the largest score dominates

(*) The proof of this and the other theorems will not be included here. They may be found in the original papers in which the theorems are presented.

everyone directly), but for all structures, even the most egalitarian It means that, whatever the structure, the uppermost man is no more than one step removed from a direct domination of every other man This is not to say, of course, that the uppermost man is not dominated by some others, but rather that, if he is, he also "dominates" them through an intermediary There are no direct and obvious consequences from this theorem for models of dominance structures, but it is such theorems as this which bit by bit allow one to know more about the properties of these structures

A second theorem proved by Landau [1953] gives the condition that a set of N numbers be a score structure Landau shows that a set of N non-negative integers $(v_0, v_1, \ldots, v_{N-1})$ is a score structure for a group of n individuals if and only if the following conditions are met

$$\sum_{i=0}^{N-1} v_i = \frac{N(N-1)}{2} \tag{4 39}$$

and

$$\sum_{i=0}^{k} v_i \geq \frac{k(k-1)}{2} \text{ for } k = 1, \ldots, N-1 \tag{4 40}$$

The first condition is obvious, since there are $N(N-1)/2$ dominance relations within a group, and each relation counts as one in the score structure, the sum of scores must be $N(N-1)/2$

Essentially, the second condition states that the sum of the scores of any k members of the group is at least equal to $k(k-1)/2$ For example, the following set of numbers is not a score structure of a group of N members even though the scores sum to $N(N-1)/2$ (7, 6, 5, 5, 2, 2, 1, 0) The last four v_i (2, 2, 1, 0) add to 5, while $4(4-1)/2 = 6$ This means, for example, that it is impossible to construct a dominance structure of eight members such that the members have the scores (7, 6, 5, 5, 2, 2, 1, 0)

Although the proof for this theorem will not be given, it is possible to see intuitively why this condition is necessary Among any k members of a group, there must be $k(k-1)/2$ relations, and each of these relations must count as part of the score of one of the k members Thus, apart from whoever else they dominate or are dominated by, this subgroup must *within itself* have this number of dominances of its members over one another Similarly, for every possible subgroup, the same condition must hold the sum of the scores must be at least equal to the number of relations which exist between these members alone

A third theorem, proved by Davis [1954], is one which is both mathematically difficult and rather important for these models. The theorem provides a way of finding the number of distinguishable dominance structures for a group of size N. This number is 2 for a 3-person group and 4 for a 4-person group, as the examples presented earlier indicate. But for larger groups the number increases rapidly, and it is this increase which makes difficult the treatment of dominance structures for more than a few members. For example, a group of size 8 has 6880 dominance structures, which means the matrix of transition probabilities must have 6880 rows and 6880 columns.

The actual theorem which gives the means of computing exactly the number of structure for a group of size N is quite complicated, and will not be given here. The procedure is based on finding the number of permutations of N integers and counting the number of cycles of various lengths in each permutation. The procedure would be difficult for large N, but for N up to around 15, it would not be too tedious.

An approximation for the number of distinguishable structures is

$$st(N) = \frac{2^X}{N!\,(N-3)!\,3} \qquad (4.41)$$

$$(\text{where } X = N(N-1)\,(N^2 - 5N + 8))$$

where $st(N)$ is the number of distinguishable structures in a group of size N.

Leeman's Sociometric Choice Model. Calvin P. Leeman [1952] has developed a dynamic model of sociometric choice and contact, which largely follows Rapoport's approach, but which is based on interpersonal choice rather than dominance relations. Leeman followed his development of this model with an experiment in which he tested the postulates of the model. Because such testing of a model has been done only infrequently among the men who have developed the models discussed here, it will be fruitful to examine this experiment and its relation to Leeman's model after presenting the model itself.

Leeman's situation is a rather simple one: he postulates that a group of people make random choices among one another, then meet each other in pairs at intervals of time, and revise their choices on the basis of these meetings. They continue meeting and revise their choices indefinitely. Leeman asks such questions as: What will be the equilibrium structure to emerge? and, what will be the probability of changing from one structure to another?

Formally, the postulates are

(A) The initial choices of the N persons are random

(B) Encounters occur between pairs of persons and only pairs Each encounter involves only one pair of persons

(C) All encounters are equally likely (Any two persons are just as likely to meet in an encounter as any other two persons)

(D) An encounter has either outcome 1 or outcome 2 for each participant in the encounter, and for each participant outcomes 1 and 2 are equally likely (He notes here that one interpretation of the two outcomes is being impressed *vs* being unimpressed)

(E) A person's choice immediately after an encounter in which he has participated depends only on that encounter, as follows

(a) If the encounter has outcome 1 for the person, he chooses the other participant in the encounter

(b) If the encounter has outcome 2 for the person, all possible choices by him are equally likely

These postulates give rise to a Markov process, in which the group structures preceding any contact constitute the states, and the transition matrix operates upon the probability distribution over states whenever a contact is made Leeman develops the model for the three man group as follows He shows that there are two possible structures (or states, as we shall call them, since they represent states in the Markov process) of the group (*)

(1, 1, 1)

(2 1, 0)

The numbers below the graphs indicate the number of choices received by each member It may easily be shown that under the assumptions of random choice (postulate A), state (1) has a probability of 1/4 of occurring, and state (2) a probability of 3/4 (†) Similarly, it may be shown that under the assumption that the outcome of an encounter is independent of the

(*) Note that while the first structure corresponds graphically to structure (1) in the three-person dominance model the second does not correspond to structure (2) in that model Nevertheless, under the random choice assumptions, the probabilities of occurrence of these two structures are the same as those that is $1/4$ and $3/4$ respectively

(†) The method of determining these probabilities is similar to that on page 96 where the probabilities of two dominance structures are computed

previous state, and depends on the outcome of the encounter as indicated in postulates D and E, the matrix of transition probabilities is

$$p = \begin{bmatrix} 3/16 & 13/16 \\ 3/16 & 13/16 \end{bmatrix} \tag{4 42}$$

That is, the probability of going from state (1) to state (2), p_{12}, is 13/16 (upper right), the probability of going from state (1) back to state (1), p_{11}, is 3/16 (upper left), and so on As postulate E indicates and this matrix shows explicitly, the probability of going to either state is independent of whichever state existed previously, being 3/16 in going to state (1) and 13/16 in going to state (2)

This information about the transition probabilities, together with the previous information about the initial probability distribution over states, allows calculation of the probabilities of being in state (1) and state (2) at any future time The matrix of transition probabilities is simply applied as an operator upon the state probabilities to show the results of an encounter The number of applications of this operator corresponds to the number of encounters

In a slightly different fashion, the equilibrium probability distribution over states can be calculated If $S_1(t)$ and $S_2(t)$ are the probabilities, respectively, that the group is in state (1) and state (2) after the t^{th} encounter then the transition matrix carries the group from one state to another as follows

$$[S_1(t-1), S_2(t-1)] \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix} = [S_1(t), S_2(t)] \tag{4 43}$$

By using this recursion equation the eqilibrium distribution, that is, $S_1(\infty)$, $S_2(\infty)$, may be found by substituting $S_i(\infty)$ for both $S_i(t)$ and $S_i(t-1)$ in equation (4 43), since by definition the probability distribution over states is constant at equilibrium. It turns out that $(S_1(\infty), S_2(\infty)) = (3/16, 13/16)$ In fact, (3/16, 13/16) is the probability distribution over states for any time after the first encounter

It is not necessary in this three person case to go through these computations since p_1 and p_2 are actually the state probabilities as well as the transition probabilities But the above directions indicate how one would go about determining $(S_1(t), \quad, S_{k_\lambda}(t))$ and $(S_1(\infty), \quad, S_{k_\lambda}(\infty))$ when λ is larger than three (k_λ is the number of possible states for a group of size λ) Leeman calculates $S_i(0)$, $S_i(1)$, $S(2)$, and $S(\infty)$ for four person

groups, where there are six possible states These calculations show that after the second encounter the distribution is very near $S(\infty)$, that is, they show that the group rapidly reaches its equilibrium state

Leeman computes one other set of properties for the three person model the mean and variance of the recurrence time for each of the two states, that is, the time it takes to get from state (1) back to state (1) The usefulness of these calculations, and in fact of any added deductions which can be made from the postulates, is in comparing the model with experimental results It may be that certain deductions from the model fit with the experimental results, while others do not If only the former deductions are used, then the model might be considered to fit the experiment when in fact it does not Leeman's experimental results provide a good illustration of this, as will be indicated shortly

The basic model as exposited above is very simple, Leeman has complicated it in a later section of his paper as follows

(1) He changes postulate C to a postulate which gives greater probability of contact to persons who are in a mutual choice relationship In a three person group, the probability of contact between a pair of persons in a mutual choice relationship is $1/3 + \delta$, and $1/3 - \delta/2$ for the other two pairs $(0 \leq \delta \leq 2/3)$

(2) Postulate E is changed so that the probability of outcome 1 (that is, definitely choosing the partner of the encounter) is greater than 5 if the two persons are in a mutual choice relationship, say $5 - \varepsilon$ $(0 \leq \varepsilon \leq 5)$

These changes provide a much more realistic situation, for they allow both greater contact between friends and a greater likelihood of choice upon an encounter, if the two are friends before the encounter Leeman presumably introduces it to show how the pure chance model can be modified to fit more closely to actual behavior (*)

Leeman's Three-person Experiment An experiment was carried out with the following general purpose in mind this model (the simple pure chance model, not the generalization mentioned above) is based on a set

(*) $S_2(\infty)$ and $S_2(t)$ are computed for this generalized model They are

$$S_2(\infty) = \frac{13}{16}\left(\frac{1}{1-\Delta}\right)$$

$$S_2(t) = \frac{13 - \Delta^t - 12\Delta^{t-1}}{16(1-\Delta)} \quad \text{where } \Delta = \frac{\varepsilon}{6} - \frac{\varepsilon^2}{12} - \frac{\delta\varepsilon}{2} - \frac{\varepsilon\delta^2}{4}$$

$S_1(t)$ of course is simply $1 - S_2(t)$

of five postulates, A through E on page 107 The experimental situation is so constructed as to fulfill postulates A, B, and C Thus, given that A, B, and C are fulfilled, a comparison of the experimental results with the model's deductions will test whether or not postulates D and E are fulfilled This general procedure is ordinarily the way an experiment is devised to test a model Certain social and psychological postulates make up the model, and the experimental conditions are set up to fulfill some of them (usually the social ones), while others (usually the psychological ones) are tested by the experiment's results

The details of the experiment are not important here It is sufficient to say that there were three persons who met as pairs in accordance with the postulates A, B, C, and after each meeting each member of the pair who met made a new choice There were 112 sequential choice patterns, and the results were as follows (*)

State 1 occurred 21 times
State 2 occurred 91 times
State 1 followed State 2, 9 times out of a possible 89
State 2 followed State 1, 6 times out of a possible 19

The estimates, therefore, of $S_1(\infty)$ and $S_2(\infty)$ (or of $S_1(t)$ and $S_2(t)$, since they equal $S_1(\infty)$ and $S_2(\infty)$) are

$$S_1(\infty) = 21/112 = 3/16 \tag{4 44}$$

$$S_2(\infty) = 91/112 = 13/16 \tag{4 45}$$

The experimental results agree perfectly with the model's prediction with respect to S_1 and S_2 However, the experimental results give as the estimate of the transition matrix

$$p = \begin{bmatrix} 13/19 & 6/19 \\ 9/89 & 80/89 \end{bmatrix} \tag{4 46}$$

which is not near the deduced matrix given in equation (4 42) Comparing these two matrices shows clearly that there was more stability in the choices than the model predicts That is, there is a considerable tendency to maintain the same choices that existed before the meeting, as the size of the diagonals indicates Leeman suggests that this indicates that the transition probabili

(*) 89 and 19 add up to only 108 trials while the actual number of trials was 112 This is because the experiment was conducted in four sessions of 28 meetings each and the last meeting in each session was considered a terminal meeting with no transition to the next session

ties should depend upon previous choices, rather than simply on the state existing before the encounter. It does appear this way at first glance, because of the strong tendency not to change which is in evidence. Such an elaboration would make the model somewhat difficult mathematically, and Leeman does not attempt to construct such a model which will fit his experiment.

If one takes another tack, however, it appears that a rather simple model may account better for the data. Rather than postulate that an individual has a certain probability of *making* a particular choice, given an encounter, it seems reasonable to postulate that he has a certain probability of *changing* his choice from what it was before the encounter. Note that this postulates a different psychological mechanism: the individual is not assumed to have his choice fully dependent on the encounter, with a particular tendency to choose the person he meets He is simply assumed to have a certain probability of changing his choice from whoever it was before to another person.(*) If a probability of change of 1/6 is assumed, then the predicted transition matrix can be shown to be:

$$p = \begin{bmatrix} 25/36 & 11/36 \\ 1/9 & 8/9 \end{bmatrix} = \begin{bmatrix} .69 & 31 \\ .11 & .89 \end{bmatrix}. \tag{4.47}$$

The experimental estimate of the transition matrix is, as before:

$$p = \begin{bmatrix} 13/19 & 6/9 \\ 9/89 & 80/89 \end{bmatrix} = \begin{bmatrix} .68 & .32 \\ .10 & .90 \end{bmatrix}. \tag{4.46}$$

Thus, such a mechanism as the suggested one gives a transition matrix very close to the one found in the experiment (†) This of course is *ex post facto*

(*) In extensive experiments with communication in small groups, MIT experimenters found that the assumption which best accounted for the data when there was a choice between alternative destinations with no reasonable basis for choice between them was a perseverance assumption (personal communication from R Duncan Luce) That is, individuals had a tendency simply to send their messages to the same group member as in the previous trial. This is the same mechanism of choice postulated here

(†) The equilibrium states posited by this revised model are $(S_1(\infty), S_2(\infty)) = (S_1(t), S_2(t)) = (1/4, 3/4)$, rather than $(3/16, 13/16)$ as the original model predicts These do not agree so well with the data for they would predict 28 and 84 rather than 21 and 91 as the number of occurrences of each pattern However, the experimental data are not significantly different from this at the 05 level, so the model is not repudiated here A model which would fit these data better is one which postulates the probability of change *away* from the interaction partner if one has already chosen him to be greater than that of the probability of change *toward* him.

theorizing, but it does suggest that this alternative mechanism, one in which the probabilities are probabilities of change of choice rather than simply probabilities of choice, may be more nearly correct than the other. Such a use of experiments, to *revise* one's prior notions about how people behave, seems at least as important as what is often their primary use, to *verify* what one already hypothesizes to be true. Another procedure which it seems should be standard practice in carrying out experiments, and which would have benefited the development of Leeman's model, is the gathering of detailed data which can test directly some of the assumptions. For example, if Leeman had tabulated the choices made after each encounter, determining whether they were made to the encounter partner or not, he would have had 224 observations by which to test directly his postulates D and E. For if the postulates are true, then about .75 of the 224 choices should have been made toward the encounter partner, .25 of them toward the third man. By testing this, postulates D and E are directly testable, thus raising a question as to why the superstructure of deductions is necessary at all. For it is only postulates D and E which are tested by the deductions. A, B, C, are predetermined by experimental conditions. If the postulates under investigation are directly testable, then why bother to add the other postulates and construct a complicated model which provides only an oblique test anyhow?

Not only could these postulates be directly tested, but by examining the patterning of the choices with an open mind about how these choices were made, it might be possible to decide on just what basis the choices are made, and let this knowledge direct one to the use of a particular stochastic mechanism. The lack of such examination in this and other models suggests what one often suspects in examining mathematical models. Many model-builders seem less concerned with the behavioral problems themselves (that is, the problem of what is an adequate mathematical representation for a given psychological or social phenomenon) than in the construction of a model for the model's sake alone. There is a certain fascination in postulating a mathematical model and deducing consequences from it, a fascination akin to the delight one feels on constructing a mechanical toy with moving parts. Model-builders often seem captivated by that fascination, forgetting that their major problem is explaining why people or aggregates of people behave as they do.

Probably few of us who have worked in this field are guiltless in this respect. Besides the fascination which a well-constructed model offers, which may distract us from the behavioral problems, there is the question

of just how the model will be of aid if it does adequately represent a behavioral phenomenon Social science is at the stage at which it does not quite know how mathematical models will help even if they are successful in mirroring behavior As a result there are seldom any great motivations toward making them do so Until we see more clearly some of the uses to which successful behavioral models will be put we are likely to continue as at present In III some suggestions about the uses of behavioral models will be made, on the basis of the examinations made here in II, but these by no means give an adequate answer

This is the extent of the examination of Leeman s model It is much like the models of status hierarchies, but it does provide two elements which those did not a) it indicates how dynamic models involving the choice relation, in contrast to the dominance relation, may be constructed, b) it indicates the relation between model and experiment, where the experiment is designed to test the model

This model, like those of Rapoport and Landau, can hardly be used in conjunction with actual groups, in a field situation In this respect it is unlike some of the static models presented earlier, which may serve as standards of measurement for certain properties of a group

Appendix 4 1.

Landau has developed a useful hierarchy index for characterizing the hierarchization of a group as a function of the score structure The index differentiates between groups in which each person dominates relatively equal numbers of others and groups in which some dominate many, while others dominate few In this appendix is presented an alternative index illustrating again (as in Appendix 3 1, p 65) a possible use of information theory in such problems as these

Measurements of information are essentially measurements of 'disorder or 'uncertainty in a system in which selection from among alternatives takes place Maximum uncertainty occurs when each alternative is selected equally often, minimum uncertainty when one alternative is selected to the exclusion of all others

Now let us think of one individual in a group, say individual i If there are N people in the group altogether, then in the establishment of a dominance structure in the group, individual i has an encounter with each of the N—1 others Each of these encounters can be considered an act of 'selection," in which there are two alternatives individual i and the other participant in the encounter The uncertainty or disorder concerning the outcome

of any single one of these encounters is obviously least when individual i either wins all or loses all. In either case, knowledge of one outcome will tell knowledge of all the other $N-2$ outcomes as well. On the other hand, uncertainty about any given outcome is maximum when the individual wins half his encounters. In such a case, knowledge about any single outcome tells very little about the other outcomes.

If we consider individual i as one alternative in a "selection" process which is repeated $N-1$ times and all the others as a second alternative, the measure of uncertainty about the outcomes is:

$$H_i = p_i \ln p_i - (1-p_i) \ln (1-p_i) \tag{4.48}$$

where $p_i =$ proportion of wins by individual i in the $N-1$ encounters and

$1-p_i =$ proportion of losses by individual i in the $N-1$ encounters.

In this way, the uncertainty of outcomes for each of the N individuals in the group can be computed. This will give N H_i's, each of which has a value dependent upon the number of wins of individual i ($i = 1, 2, \ldots, N$). If a group of size three has a score structure (1, 1, 1), this means that each individual has won one of the two encounters in which he has engaged. The H_i's would thus be

$$H_1 = H_2 = H_3 = -\tfrac{1}{2} \ln \tfrac{1}{2} - \tfrac{1}{2} \ln \tfrac{1}{2} = -\ln \tfrac{1}{2}. \tag{4.49}$$

If, on the other hand, the score structure is (2,1,0), then

$$H_1 = -1 \ln 1 - 0 \ln 0 = 0 \tag{4.50}$$

$$H_2 = -\tfrac{1}{2} \ln \tfrac{1}{2} - \tfrac{1}{2} \ln \tfrac{1}{2} = -\ln \tfrac{1}{2} \tag{4.51}$$

$$H_3 = -0 \ln 0 - 1 \ln 1 = 0. \tag{4.52}$$

The uncertainty or disorder in the system as a whole would seem to be simply the sum of these H_i's. However, each encounter is counted twice — once for each of the two persons playing a part in it. Thus the uncertainty measure for the system as a whole is

$$H = \frac{\sum_{i=1}^{N} H_i}{2} \tag{4.53}$$

or

$$H = -\sum_{i=1}^{N} p_i \ln p_i. \tag{4.54}$$

This, then, is the alternative 'hierarchy index' which information theory suggests. It has its maximum when the score structure exhibits perfect equality (i.e., maximum uncertainty), and its minimum when the score structure exhibits a perfect hierarchy (i.e., minimum uncertainty). It thus varies inversely with the degree of hierarchization; for heuristic aid, the inverse of H, or alternatively, $H_{max} - H$, might be a more desirable index. Both of these would be maximum for the perfect hierarchy, minimum for perfect equality.

This formula for H looks like the usual equation for a measure of the disorder of a system with N alternatives. However, it must be remembered that the structure of this 'selection' process is different from one in which selection is directly from N alternatives. The major consequence of this difference is that the p_i's in this formula do *not* sum to 1.0, as in the usual situation, rather they sum to $\frac{N}{2}$. This means that for $N > 2$, the minimum disorder (perfect hierarchy) cannot be zero, but some value greater than that, increasing as N increases. In the example of a 3 person group given above, where the perfect hierarchy is (2, 1, 0), H is $-\ln \frac{1}{2}$, rather than zero. The reason for this is simply that in a group larger than two, some persons must both win and lose, only two persons at most (the top and bottom man) can win or lose all encounters.

This increase of H_{min} with N seems to correspond to the kind of index one might want, as N increases it is certainly true that in a perfect hierarchy, the number of intermediate positions (persons who have lost some and won some encounters) increases so that there are many more *levels* of position in a large perfectly hierarchized group than in a small one. It might be desirable to incorporate this objective difference into one's index, or it might not be.

This simple example is introduced to illustrate again the kind of uses to which measures of information, or disorder, can be put (*). In structural problems like those which have been examined in this section, there is often need for measures of the degree of 'order' exhibited by the structure (†). Information theory provides a measure which is derived from explicit as-

(*) It can be shown, though it will not be done here, that this index and Landau's are logically related. Landau's index and H are both functions of the variance of the v_i (where $v_i = (n-1) p_i$).

(†) A number of measures designed to isolate various tendencies or properties of a social structure are presented in Nehnevajsa [1955]. See also Coleman [forthcoming] chapter on Structural Sociology.

sumptions about the structure, to the degree that these assumptions are accepted as reasonable, the measure is less arbitrary than many of the *ad hoc* indices commonly used in social research

5 MODELS OF GROUP ACTION

David Hays and Robert Bush [1954] have developed a model to characterize group learning in a situation where the group is repetitively faced with a choice between two alternatives The situation is analogous to a learning situation for individuals, in which one alternative is more likely to be correct, and as time goes on, the individual becomes more and more likely to choose this more probable alternative For the analogous "group learning" situation, Hays and Bush develop two alternative models based on different assumptions about the influence of individual members' choices upon the group's choice (*)

A quite different kind of model, but also one which attempts to account for group problem solving behavior, is developed by Irving Lorge and Herbert Solomon [1955] Lorge and Solomon use two alternative models to account for the results of a classic experiment which compared the ability of groups and individuals to solve eureka-type problems These results had been interpreted to indicate that people performed better on such problems when they were in groups than when alone Lorge and Solomon show that simple assumptions can account for the apparently better results of the group without postulating any effect due to the group context itself

These two approaches to model construction are similar in that they attempt to account for some single unitary action of the group as a whole In the Hays Bush case, the action is choice between two alternatives, in the Lorge Solomon case, it is solution of a eureka-type problem The two models are not at all alike mathematically, but have only this substantive similarity Yet in a second way as well they are similar they each postulate alternative models, and use the experimental results to select between the two alternatives Because of these similarities, the two models will be examined in this one section, though neither will be treated exhaustively

The Hays-Bush Stochastic Model of Group Learning. Bush and Mosteller, among others, have developed stochastic models to account for individual learning behavior [1955] In the simplest situation the subject is faced

(*) V B Cervin [1957] has very recently developed a model for processes of persuasion in groups, using reinforcement theory His model appears to be a promising direction of work.

at each trial with the possibility of either of two events (such as the flashing of a light or its failure to flash) occurring, on the basis of his experience from preceding trials, he must choose which of these events will occur on the next trial Ordinarily the experiment is so set up that one event is somewhat more probable than the other For example, in some experiments, the probability of the light's flashing (event 1) on any trial is 75, and the probability of its not flashing (event 2) is 25

The stochastic learning model of Bush and Mosteller for this situation postulates that the occurrence of event 1 on one trial decreases the subject's probability of choosing event 2 on the next trial by some constant proportion That is, if q_n is the probability of choosing alternative 2 on trial n, then when event 1 occurs on trial n,

$$q_{n+1} = a\, q_n \tag{5 1}$$

where a, varying between zero and one, is the proportion by which q is reduced A similar equation holds in the case that event 2 occurs on trial n, the probability of choosing event 1 on trial $n + 1$ is reduced by the same proportion, a, from p_n If a is one, then there is no change in the probability of choice, if a is zero, then the subject always chooses the event which occurred on the last trial If a is somewhere between these two extremes, then the subject chooses less and less often the event which occurs infrequently until finally some equilibrium value for q (or alternatively, p) is reached

This model has been shown to describe adequately an individual s choice behavior in the situation described above But what happens when a group of two, three, or more persons is instructed to make a single choice as a group on each trial, just as the individual is usually instructed to make his choice? Hays and Bush suggest two extreme alternative hypotheses 1) the group acts like a single individual, so that the model which describes the single individual s behavior will describe without change the group's behavior, and 2) the individual members make choices independently, as described by the individual model, and then these choices are combined by some suitable voting procedure to give the group's choice In a group of three members, as Hays and Bush used in experiments, the hypothesized voting procedure was simply a majority decision whichever alternative was favored by two or three members was the alternative chosen by the group

It is important to remember that the groups were not instructed to behave in these ways, rather, there were no instructions about how to reach a

decision. These two alternative models represent differing hypotheses about the way in which the decision was reached.

In the first model (which they call the "group actor" model), the Bush-Mosteller procedures can be directly used for fitting the model to the data. Twenty-one three-man groups were used in the experiment, each tested for 100 trials. From the resulting data, and with .75 as the probability of occurrence of event 1 throughout these 100 trials, two parameters are computed: p_o and a. With these two parameters, it is possible to calculate the expected probability of choice at each trial.

The voting model requires a modification of the basic Bush-Mosteller equations. If the probability of choice of alternative 1 at trial n is p_n for the Bush-Mosteller model, the group's choice is as follows: assuming that p_n is the same for all three group members, the probability of group choice p_n is Pr [all three choosing alternative 1] + Pr [two choosing alternative 1 and one choosing alternative. 2], or $p_n^3 + 3p_n^2\ (1 - p_n)$. With this as a starting point, the Bush-Mosteller model can be used to establish equations for this voting model, to solve for the two parameters, p_o and a, and to find the expected probability of choice at any trial.

The equations by which these results are found will not be reproduced here; they are complicated and not essential for understanding the intent of the model and related experiments. But the graph which shows the actual results together with the fit of the two models is shown below:

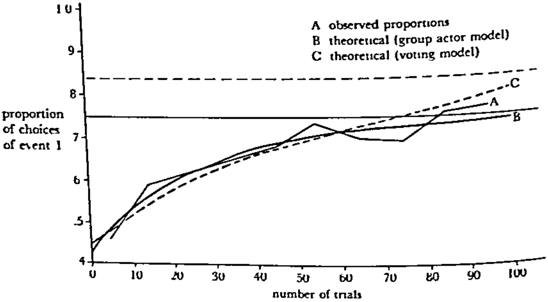


Fig. 5.1*

(*) At intervals of ten trials are plotted the proportion of choices of event 1 by twenty groups for the block of ten trials. The horizontal dashed line is the asymptote for the voting model, and the solid line is the asymptote for the group actor model.

As the graph suggests, the results lie squarely between the two models Neither model can be rejected by the statistical tests applied by the author, and in fact, each provides about the same goodness of fit to the data

The results, then, are inconclusive As the authors point out a test might have been possible if the groups had been carried beyond their 100 trials According to the group actor model, the groups had already reached their asymptote (at 75), while the voting model indicates that they are still some trials away from the asymptote

Because the data lie between the group actor model and the voting model, the authors consider the possibility of some intermediate model This opens up a whole range of models, for, as it turns out, the group actor model is equivalent to assuming that one person decides and the other two members are fully influenced by him, the voting model assumes that each member acts independently, after which a vote is taken A model which incorporates *some* degree of support is intermediate between these two extremes, the authors propose a range of such models, and a support parameter to determine which of the models is operative

Hays and Bush indicate that the work they report is only a starting point for interaction models, and suggest that in future work the interaction should be more rigidly controlled Their experiments indicated that inter- action among the group members varied greatly for different groups with obvious consequences for the resulting group choices For example, they report that most of the groups laid out a conscious voting plan at the be ginning of the experiment, though not all groups adhered to their plan throughout the experiment Since it was precisely the nature of the inter- action which was under investigation (*i e*, 'follow the leader" *vs* majority rule), such conscious plans would seem to determine *a priori* the very matters which were intended to be the result of unplanned relationships arising from the interaction However, this might not be so It could be that the unplanned and unanticipated relationships would completely override the intended patterns If this were so, then such a result would be quite im portant

Parenthetically, it might be noted that, since the major focus of the model is upon this decision making process, it might have been simpler not to carry along the excess baggage of a stochastic learning model, but instead focus directly on the decision process itself For since the learning model is already confirmed for individuals, then the only part of the group learning model which remains to be confirmed is the group decision process whether the group makes its choice in a follow the leader fashion, or by voting If one really desires to investigate this question, why not do it under circum

stances in which the data test more directly which of these two procedures — if either — the group uses?

A possible answer to this question is that the stochastic learning situation provides the kind of setting within which "natural" or "unconscious" or "fundamental" modes of group decision-making occur. As a consequence, the model will detect in a way that a more direct experiment would not, what is the nature of this "fundamental" decision-making process.

Whether or not this is the authors' answer, it is nevertheless true that a clear-cut situation is constructed: two alternative models are developed, with alternative predictions about the resulting group choice, and with an experiment to differentiate between them.

One of the authors [Hays, 1956] has extended these initial experiments by restricting the interaction among participants so that their only "interaction" is in seeing the choice made by the other person; by having each subject, rather than the groups as a unit, make choices; and by posing a reward at the end of each trial. However, to date these further experiments have not been amenable to a simple mathematical model. Because of this, these further experiments will not be discussed here. However, one interesting tentative result has been found: in such a situation, pairs of persons in interaction develop "patterns" of responses which are peculiar to them as a pair but lead to a high proportion of rewards, because they mutually learn to anticipate what the other will do. The pattern is not something determined by the reward structure, for it differs from pair to pair. It is a "norm" which develops as it leads to mutually rewarding outcomes. This result is reminiscent in many ways of Sherif's autokinetic experiments, and it may be that the theoretical relationship between the two is a close one.

This model will be mentioned again in III, where questions are raised regarding the direction in which various work is leading. Any questions about the purpose and goals of this model-building will be deferred until that point.

The Lorge-Solomon Group Problem-solving Model. In 1932, a set of experiments were carried out in which groups and individuals were compared with respect to their ability to solve certain mental puzzles (Shaw, [1932]). The results, which have generally tended to be confirmed throughout the two decades since Shaw's time (Taylor and Faust [1952]), showed fairly conclusively that groups performed better than individuals. Though one recent experiment (Moore and Anderson [1954]) has shown groups and individuals to be alike in their ability to solve certain problems

(except that groups showed a lower *variability* in successful performance), the earlier result of Shaw has been generally accepted The better performance of the group has been attributed to the effect of group interaction, both by Shaw and by later writers Lorge and Solomon accept the results but raise a question of interpretation is it really necessary to postulate that group interaction — or any other unanalyzed "group effect — causes the differences found? Might not the results be explained simply by the added number of individuals available in the group, so that without any interaction at all and assuming the average level of ability was the same, the groups would show better performance? Or if this explanation does not suffice, suggest Lorge and Solomon, perhaps a slightly more complex one will the solution proceeds in stages, and there is no "interaction effect" or any other group effect within each stage, however, if one member solves one stage of the problem, then the group is freed to work on the second stage Thus in the extreme case, suppose there were two kinds of individuals those who could solve the first stage of a problem but not the second, and those who could solve the second, but not the first Then no individuals could solve the problem alone, but any group with at least one person of each kind in it could solve the problem

With this as their general orientation, Lorge and Solomon reanalyze Shaw's data, using first the single stage model, then a multi-stage model assuming all individuals to be alike, and finally a two stage model assuming complementarity of ability among group members The problems Shaw used in her experiment are classic problems of transport, such as the problem of how three jealous husbands and their wives can all be transported across the river in a boat which will carry only three at a time, with only husbands doing the rowing, and no woman allowed in the presence of another man unless her husband is also present Three individuals out of twenty-one solved this problem, while three (four membered) groups out of five solved it Thus on the surface, the groups are far better than individuals, since only 3/21, or 14%, of individuals solved the problem, and 3/5, or 60%, of groups solved it

But perhaps the data can be viewed a little differently Suppose each individual has a probability P_I of solving the problem Then, supposing no interaction at all, the group probability of solving the problem is as follows

$$Pr\left\{\begin{matrix}\text{group of } k \text{ members}\\ \text{solve problem}\end{matrix}\right\} = 1 - Pr\left\{\begin{matrix}\text{no member solves}\\ \text{problem}\end{matrix}\right\} \qquad (5\ 2)$$

$$= 1 - Pr\left\{\begin{matrix}\text{member}\\ 1 \text{ fails}\end{matrix}\right\} \quad Pr\left\{\begin{matrix}\text{member}\\ k \text{ fails}\end{matrix}\right\} \qquad (5\ 2\text{a})$$

$$= 1 - (1 - P_I)\,(1 - P_I) \quad (1 - P_I) \tag{5 3}$$

$$= 1 - (1 - P_I)^k \tag{5 3a}$$

The best estimate of P_I from the data on individual performance is 14, since 14% of the individuals solved the problem Substituting this value in equation (5 3a),

$$Pr\left\{\begin{matrix}\text{group of four}\\ \text{members solves}\\ \text{problem}\end{matrix}\right\} = 1 - (1 - \ 14)^4 \tag{5 4}$$

$$= 46 \tag{5 4a}$$

This value of 46 compares with the actual value for groups of 60 Thus the simple one stage model does not fit, nevertheless, the use of this simple model shows that groups and individuals are much more alike than the raw data indicated Or, to put it differently, a great deal of the apparent superiority of groups (60 compared to 14) can be accounted for simply by the individual abilities, without postulating *any* sort of effect of group membership

A more complicated model supposes that there are multiple stages to the solution of the problem, and that an individual or group may falter at any stage in the solution In this case, the fact that people are in a group allows one person to build upon another's work, without himself solving the first stage, he can contribute to the solution by beginning with the solution to the first stage and going on to solve the later stages

With this assumption, the relation between group and individual probabilities of solution is found as follows

$$P_I = P_{I_1} \quad P_{I_2}, \qquad , P_{I_s} \tag{5 5}$$

where P_{I_i} is the individual's probability of solution of stage i, and s is the number of stages Then

$$P_G = [1 - (1 - P_I)^k]\,[1 - (1 - P_{I_2})^k], \qquad , [1 - (1 - P_{I_s})^k] \tag{5 6}$$

where P_G is the probability of group solution An analysis of this equation shows, as did equation (5 3), that simply on logical grounds alone, with no psychological differences at all due to group membership, the group will more likely solve the problem than will the individual, if the problem is of this multi stage form And an examination of equation (5 6) would show that

model, just as it was not possible to test the previous multi-stage model. The logic of this model remains clear, however, and it is similar to that underlying the previous multi-stage model: The group may perform better because its members can build on one another's work, sometimes having complementary skills. The lone individual, in contrast, is kept from solution if he lacks the ability to solve any single stage of the problem.

This model will not be developed in full here; the above comments should suffice to indicate its intent. Further remarks concerning Lorge and Solomon's models will be made in III.

III

CONCLUSION

6 SUMMARY AND EVALUATION

Introduction Probably the most remarkable fact about these approaches to mathematical social science is their wide diversity What is clearly evident is that the various model builders have begun with quite different purposes in mind While this is obvious upon examination of these models, it is a fact which is often not recognized We often forget that mathematical models are used for diverse purposes and cannot at all be judged from a single point of view Stephan, developing a generalization about relative participation rates on the basis of quantitative data, certainly has a different goal in mind than does Simon, formalizing the qualitative and discursive hypotheses of Homans Rapoport's purpose in developing a model of status hierarchies is no less different

Because this diversity of purpose is one of the most important generalizations to derive from the juxtaposition of small group models, part of this section will be devoted to an examination of these purposes and their implications for the future of theory and research in this area There will be no attempt to infer what the various motivations were in each case of model building, but rather a concentration on the fruits which might legitimately be expected from each approach, and the general direction in which each appears to lead In summary, the several approaches are

1 *Formalization of* existing verbal propositions which relate a set of interdependent variables, and which specify only the direction of the relation (*e g*," increases with an increase in) The example of this is Simon's formalization of Homans' hypotheses about sentiment, interaction, and activity in a group

2 *Development of* a quantitative relationship between various attributes of a group (or its members) based on generalizations from quantitative data The examples of this were the models based on relative rates of participation among members of a discussion group as developed by Stephan, Bales, and others

3 Models based on simple, all-or-none binary relations, which ordinarily correspond to sociometric relations, communication relations, or dominance relations These models were of several kinds

a Random models, in which an expected network of relations is derived from the assumption of random outcome of choice, communication, or dominance The expected numbers of cycles of various sizes, of isolates, of chains of a given length, etc were calculated by authors who used this approach

b Modification of these models, which involved changing the assumption of randomness in one direction or another to characterize some tendency (*e g* , toward clique formation, toward mutual choice, toward choosing "upward,' toward status hierarchization)

c Models which essentially give a mathematical machinery for tracing out certain properties of a network, *e g* , cycles of various sizes, if one knows the relational structure in matrix form

d Models which introduce a mechanism of change so that an initial structure could be modified through time by meetings and subsequent changes in single relations

4 Models of group action, which were developed to account for, or predict, the behavior of groups in some simple situation These models were of two kinds

a A stochastic learning model designed to account for "group learning' in a simple choice situation constructed around a probabilistic occurrence (*e g* , a light which flashes with a probability of 75) The model discussed was that of Hays and Bush

b Models designed to account for a group's problem solving ability in terms of the abilities of individuals and the way they combine to solve the problem. Lorge and Solomon developed the models which were discussed here

These approaches will be examined in order, in each case attempting to see in what direction the work is leading, and speculating on the immediate and ultimate fruitfulness of the approach

Formalization of Interrelated Qualitative Propositions. The translations of Homans' and Festinger's qualitative hypotheses into mathematical models represent the most ambitious attempts yet to formalize existing verbal theory in sociology or social psychology There has been much discussion about formalizing existing theories, from grand scale systems like that of Marx to the much more modest sets of propositions like

those of Homans and Festinger Almost invariably, these verbal theories in social science have the following properties they specify a set of relevant variables (though they seldom give prescriptions for measurement of the variables), they specify the causal relations existing between these variables, and they indicate the direction of variation (*e g*, '*the more Y, the more X*') Beyond this they seldom go, limiting the theory to qualitative rather than quantitative statements

Despite much discussion there has been little actual work in formalizing such theories, so until now it has been difficult to know what the fruits of such formalization might be The papers by Simon, and Simon and Guetzkow, allow for the first time an evaluation of the possibilities of this approach

In one sense, both Homans and Festinger carry out a formalization themselves They each take generalizations based on a large body of data and develop a set of propositions using a limited number of variables They state these propositions qualitatively, but rather precisely and unambiguously The propositions consequently represent probably the limit to which one can go in the verbal development of such qualitative theories

What was done beyond this in Simon's mathematization? Primarily two things First, a translation from words to mathematical symbols, simply a restatement in another language, and second, making deductions from the translated propositions, using tools of mathematics These seem rather simple and innocuous changes introducing nothing which should lead to controversy, and adding little of value Yet attempts like this have both led to controversy and have added value to the verbal statements What are some of the sources of controversy and of added value?

Probably the most serious charge that has been leveled at these formalizations is that they are based on a superficial understanding of the hypotheses themselves Homans feels the formalization, while consistent with his hypotheses, does not capture all that is said in them (Simon [1952], p 204, fn 4) Festinger has expressed similar beliefs about the formalization of his hypotheses

Leaving aside the question of just how accurately and fully the substance of these hypotheses is reflected in these specific cases of mathematical translation, it should be recognized that a fundamental problem does exist If one attempts to develop a set of interrelated equations from which deductions can be made, these requirements must be met 1) All the suggestive and richly descriptive aids to the imagination which exist in the verbal discussion must be shed, reducing the model to a skeleton which may be so bare that the resulting model may seem sociologically trivial, 2) the number

of variables must be kept to a minimum, so that the postulates become more than just a set of equations with a great number of exogenous variables, and few endogenous ones That is, variables which act as the "independent" variable in one relation must in turn be indirectly affected by the dependent variable of that same relation Unless this requirement is met, the propositions do not form a "system", they remain a series of discrete propositions, each with its own independent and dependent variables, but with no interrelation between the propositions 3) In order to make deductions, using either the geometric method presented in II or the analytical method presented in Appendix 2 2, all but two of the equations must be algebraic, and the other two must be differential equations The system must ultimately be reducible to two simultaneous differential equations There is no limitation of this sort ideally, but because the mathematical methods of deduction do not exist, these requirements are necessary (*)

These requirements severely restrict the kind of propositions which are amenable to treatment in this fashion They mean that whatever the situation described by the verbal postulates, it must be forced into this frame work, which may be artificial For example, a very rigorous translation of Festinger's hypotheses would introduce numerous other variables, differentiating between perceived deviation of opinion and actual deviation, and letting some of the variables refer to the individual rather than to the group as an aggregate

Whether such simplifications as a model builder must make in a case like this are justified is a question which can only be decided in individual cases But it depends in part on one's aims is it better to sacrifice some richness of meaning in order to investigate fully the interrelations between the variables, or is it better to keep the richness but leave unexplored the dynamics of the system and the effects of various feedbacks? For *elaborating* a theory to encompass a wider range of behavior, it is probably best to keep the richness, together with all its suggestiveness for various avenues of elaboration For *testing* the theory, it is probably best to simplify to the extent necessary for exploring all the feedbacks of the system Otherwise, as Simon and Guetzkow show, supposed tests may not at all test the hypotheses for which they are designed

Probably one reason for the resistance among social scientists against formalizations like the two discussed here is that social scientists are more

(*) One way of treating this problem suggested by Simon is to single out the two *slower-acting* processes to be treated as differential equations while the processes which reach an equilibrium more quickly can be represented simply by equations of state, *i e*, algebraic equations

concerned with elaborating theory than with testing it Theories like Homans' and Festinger's are believed to be true by many social scientists on the basis of introspection and experience, yet are not believed to be true in any ultimate sense That is, while these propositions are conceded to be generally true, questions arise about their future usefulness are the particular concepts and relations in terms of which the generalizations are stated a fruitful way of organizing the phenomena under consideration? It is at this point that many social scientists balk, and instead of taking these theories seriously, as a theoretical underpinning for their work, use them only as generalizations which can help in finding the right way to organize the same phenomena The question of just how such theories are best used, and what really is their value, is a fundamental one, and, as the discussion above suggests, one which has important implications for their formalization into a mathematical model

This question cannot be answered here, however, an analogy may give some insight into the answer Suppose one were investigating the motion of metal balls, and suppose that there was no theory of mechanics, or even a concept of weight Then it would be possible to develop numerous qualitative generalizations on the basis of observation, somewhat as follows

"A ball will roll down a board if the board is tilted up
'The smoother the board and the ball, the faster it will roll '
'The faster the ball is rolling when it reaches level ground the farther it will roll '
'The higher the board is tilted, the faster the ball will roll'
'It takes more work to raise the board and ball, the steeper the angle to which it is raised

These and many more such qualitative propositions could be derived from observation It is obvious, however, that they do not constitute a 'theory which can serve as the basis for elaborate deductions But what role *would* they play in the development of the science of mechanics? Probably this along with numerous other such propositions, they would give more and more insight into the underlying processes at work, and the variables by which the ball s behavior could best be described In other words, they would stand as evidence which would aid in developing an explanatory theory, once such a theory were developed, this quantitative evidence would all be deducible from it

It may be that most of our qualitative generalizations in social science will have a similar function, rather than serving as the postulates of a deduc-

tive system, they may serve as evidence from which an explanatory theory can be inferred, and as deductions once that theory is inferred

Whether a role like this or the opposite kind of role Simon has given to these qualitative propositions will turn out to be the correct one probably hinges on the *variables* involved in these models if these variables, when refined, turn out to have the status that mass has in the theory of mechanics, that is one thing, if they turn out to be only roughly correlated with the fruitful variables, as 'size' of an object is correlated with mass, this is quite a different thing

But putting aside this fundamental problem, what are some of the limitations and values of a formalization like Simon's? Perhaps the most obvious limitation of all is that it cannot be used to give quantitative predictions about group behavior This is of course inherent in the nature of such models based on qualitative propositions, they can give no more than qualitative predictions To some who equate numbers with mathematics this means they are not really "mathematical", the models will be a disappointment by not giving numerical predictions Certainly this is a limitation on a model's usefulness, but it is simply a reflection of the limitations of the qualitative verbal propositions on which the model is based

Another limitation of such models relates to the measurement problem, which was discussed in detail in the examination of the Simon Homans model There the conclusion was reached that the variables of that model were so vaguely specified that the model's predictions would hardly be valid That is, without some directions concerning measurement of the variables, the measurements themselves are likely to differ so widely as to disrupt the model's predictions In the model constructed from Festinger's hypotheses, the problem is even more acute, for the variables of that model are more difficult to measure than Homan's friendliness, interaction, and activity Thus unless some solution to this problem of measurement is made, it appears that the accuracy of the predictions of the model is contingent upon the particular measurement whims of the investigator engaged in testing the model The problem could be partially solved by one or another method, one partial solution is suggested by Simon and Guetzkow They suggest that the model's scope be limited to those situations in which each member of the group changes in the same direction on each variable from time 1 to time 2 This solves the aggregation problem, since under this condition any two measurements on the individual level which are monotonically related will give aggregate variables which are monotically related But this is only one aspect of the problem the more general difficulties of measurement of such variables, discussed in II, remain.

Both these last two limitations and some of the difficulties raised previously are really questions not about the formalization, but about the role of qualitative propositions or generalizations which form the basis of much social science If we take the usefulness of such propositions as given, and if we assume that at least one part of their role is to trace out the causal structure, positing particular concepts and relations, then the following kinds of values appear to arise from a formalization

1 A major value of formalization lies in bringing assumptions and problems out into the open Some of these are

a The formalization makes one face the question of whether a particular verbal statement ('*X increases with increase in Y*') is meant to represent co variation or a structural relation That is, does a statement that X varies directly with *Y* mean that both *X* and *Y* are affected in the same direction by changes in a third variable, or does it mean that an increase in *Y* causes an increase in *X*? The ambiguity of words allows the statement to be interpreted in either way, and it is only upon formalization that one is required to make clear what is meant

b It removes other sorts of ambiguity from the postulates The very fact that it is possible to make alternative formalizations of a set of verbal postulates (thus leading to different predictions) indicates that the words carry much ambiguity Resulting disagreements about which formalization is correct are nothing more than latent differences in interpretation brought into the open by the formalizations Once out in the open, these differences may be resolved by empirical test, and knowledge is thereby gained

Both the above values of a formalization derive simply from the translation into mathematics, without the operation of drawing deductions Thus if one wants primarily to clarify a set of relations in the ways suggested above, a model can be developed with no intent to draw deductions This removes many of the constraints mentioned earlier which a formalization like Simon's must meet The translated postulates in such a model could be elaborated and enriched as much as desired to reflect the complexity of the situation, so long as no deductions are desired

2 However, the added values arising when one does make deductions may be important as well Some of these are

a The lack of independence of postulates or inconsistency between postulates is made apparent by relating the formalized postulates with one another For example Homans and Festinger each introduce as an independent postulate one which is deducible from the others

b The formalization brings out the various feedbacks in the system, and

by so doing aids in the design of experiments, showing just how certain postulates can be tested, and indicating what experiments will fail to test them For example, Simon and Guetzkow show that an experiment such as Kurt Back's, which is used as evidence for some of Festinger's hypotheses, could test them much better if the formal model had been used to aid in designing the experiment They also show that the Festinger, Schacter, and Back housing experiment tests the set of hypotheses in quite a different fashion from the way that Festinger suggests when he uses this experiment as evidence for one hypothesis

c One of the most severe limitations on verbal theories which form the bulk of social science is that one can go just so far and no further, we are just not able to handle mentally more than a certain number of relations (which are often ambiguous to begin with) We simply are unable to see all the implications of related hypotheses, and we thus limit our investigations to the most proximate of the implications This may be an important stumbling block in social science, if so, a formalization of some of the verbal theories which exist at present might allow a break through in certain aspects of theory Whether such major gains as this would actually accrue or not, it is nevertheless true that a formalization such as the one examined allows deductions to be made which are far from obvious upon a careful examination of the verbal hypotheses

3 Finally, there is at least one important gain which does not occur until several competing theories in the same area are formalized Formalization should allow rigorous comparison of theories in place of vague disagreements which are never settled It should be possible in many cases to design crucial tests which would give results confirming one of the theories and disconfirming the other In general, the specific areas of disagreement between the theories could be brought into the open, with all the benefits which would arise from this

But I have written above in terms of ' should", this gain from formalization of theories is only a potential one, for such formalization of competing verbal theory or sets of propositions simply does not exist Homans' and Festinger's hypotheses, for example, deal with subject matters just slightly too far apart to allow comparison between models developed from them

Before leaving the examination of limitations and values of the approach to model building represented by Simon's and Simon and Guetzkow's work, one further point should be mentioned There may be a distinct "psychological" disadvantage to such an approach as this Most social scientists who develop verbal theories are men whose imagination and creative ability

outrun their mathematical skill Even where this is not so, they probably prefer to relax mathematical precision in developing theory or broad generalizations of a qualitative nature Given the general aura of prestige surrounding the use of mathematics in social science — at least in some circles — substantive investigators may be inhibited from subjecting their early results and generalizations, suggestive and stimulating though they may be, to the cold scrutiny of mathematics This is probably particularly the case when such scrutiny requires that some of the richness of the results be sacrificed If a substantive investigator is bedeviled by the feeling that any attempts of his to state his results with precision and generality will be pounced upon by the every-ready formalizer, he may be less eager to make the attempt This is extremely unfortunate, for formalization of a theory or set of propositions is a second step depending completely on the creative step of building the verbal theory Whether such theories are perfect or not is less important than that they exist at all

In contrast, a psychological advantage of quite a different sort may begin to accrue from such formalizations It begins to set men *thinking* in terms of differential equations and the like, rather than discursive and ambiguous words It may easily be that the major contribution of today's mathematical models to tomorrow's social science will be that they began some social scientists thinking in mathematical terms

Development of Quantitative Generalizations Concerning Behavior in a Group. The work examined in Section 3 is certainly a different approach from the one just discussed Stephan and the others who have worked in this area are concerned with *quantitative generalizations*, not with *qualitative systems* There is nothing dynamic in the 'models they have developed Instead, there are simply equations which account for a great deal of data using only a few parameters

This approach is at once more modest and more ambitious than the "qualitative system' approach discussed above It attempts to account for only a single kind of behavior — relative rates of participation among group members — rather than an interlocking system of dispositions, behavior, and external conditions Thus it is in one sense a modest approach But it attempts to account for this behavior quite precisely, and in this sense it is an ambitious one There have been extremely few generalizations in social research which have any claim to quantitative precision, and it would indeed be an accomplishment if quantitative generalizations which were non-trivial and valid over a reasonably wide range of groups were de-

veloped The data and the descriptive models of Stephan and others may be a beginning in this direction

Perhaps the greatest immediate value of such work as this is the stimulus it provides for further work Faced with such persistent regularities, and the descriptive model which summarizes them, quantitatively-oriented social scientists have a real problem with which to work Since the descriptive model is not an explanatory one (that is, the mathematics does not reflect some sort of social or psychological process which might give rise to these data, as it might if it were a stochastic model, for example), then the regularity remains a puzzling one, and offers a problem for explanation

As the model stands, its usefulness is not great, for the regularity is a statistical one, not meant to hold for any single group meeting But once an adequate explanation is offered, then this explanation itself might make some real advances in our knowledge about behavior in groups For example, an explanatory model might indicate that once an ordering is established on the basis of frequency of interaction, then individuals have a certain probability of replying to someone above them in the order, and a lower probability of replying to someone below them in the order This would then tell something about the dynamics of discussions in groups, a matter about which little systematic knowledge exists In any case, an adequate explanation of the data would give quite useful information about discussion groups, for it would postulate certain behavior tendencies on the part of group members, and from this allow deductions corresponding to the data to be derived These behavior tendencies would essentially be hypotheses about the way people behave in group discussions

Just what is the next step in the development of this approach? That is, given the generalization and the descriptive model which summarizes it, what are the paths which would make this result a more useful one both for the further development of theory and for immediate practical uses? Probably the most pressing need is for more data to test various possible explanations of the regularity, and to give further insights into the processes which are occurring For example, in any attempt to build a stochastic model to account for the regularity, one important kind of data is relative rates at different points of time in the session If the data as presented by Stephan were split, to give relative rates for the first half and the second half of each session separately, the question about equilibrium and when it becomes established could be answered

More generally, there are a number of possible hypotheses — like those which Stephan and Mischler test in their paper — which could be tested by

gathering more detailed data Testing of these hypotheses should sooner or later lead to the development of an adequate model of the process involved

This approach of Stephan and others to the development of a mathema tical model is in direct contrast to the approach which is most in vogue at present in mathematical social science That is, much mathematical model-building begins by setting up postulates which are not based on experiments or experience, but are merely reasonable, and constructing from them a model (*) Deductions are drawn, then there is an attempt to test the deduc tions through experiment or observation, but ordinarily experiment Of the models examined in this paper, Leeman's, Hays and Bush's and to a lesser extent Rapoport's and Landau's dynamic status hierarchy models exemplify this approach In these models, postulates are set up, deductions derived, and the deductions compared with (a) experimental results (in the Leeman and Hays Bush case) or (b) data gathered from non experi mental observation (in the Rapoport and Landau case) The approach may be loosely characterized as deductive, that is, setting up *a priori* postulates, developing a model, then testing its validity These models will be discussed in more detail below, but these comments suffice to illustrate the difference in approach In contrast, the approach of Stephan and others may be loosely characterized as inductive, that is, beginning with an empirical result and investigating further until the underlying process which produces it is discovered

Relational Models *The relational models represent no clear 'philosophy of model building* as do the first two approaches, simply because there are a number of different approaches under this heading However, it may be useful to comment upon the one element they have in common the investi gation of structures composed of all or none pairwise relations This area of investigation holds a special appeal to the mathematician or mathema tically inclined social scientist, for at least two reasons

First, it presents him with numerical data, which he can feel are not so subject to the vicissitudes of measurement as are many numerical data in social science (†) That is, even though two different sociometric questions

(*) I do not mean to include here those models which are explicitly based on the prin ciple of rational behavior such as game theory Th s model serves the important purpose of showing impl cations of a set of postulates which correspond to what we think of as rational behavior Therefore the model at least can serve as a normative one telling what is best to do in various circumstances given certain assumptions about the opponent s strategy

(†) This is not to say that different methods of measurement will not give d fferent group structures See for example Helen H Jennings [1950] in which two different types of sociometric questions (living with and playing with) elicited not only different re sponses from individuals but quite different group structures

may generate different structures, a structure based on one question has its own particular meaning, and is ordinarily not a useless artifact of some ill designed measurement procedure What we generally object to in social science measurement is the fact that the measurement process often sets up equivalence classes of questionable validity To give an example from a different area of measurement, a certain method of attitude scaling (the Likert technique) has multiple response categories (varying from something like "very strongly agree' to "very strongly disagree') These are weighted from +2 to —2, and the scores from several questions added to give a total score This then establishes an equivalence class between, for example, responses of (+2, —2, +2, —2) and (0, 0, 0, 0) in a four-question scale Both of these sum to zero, and individuals with these different response patterns would be classed together Clearly, individuals giving these two different patterns of response have different attitudes, not equivalent ones (*) Relational models largely by pass this difficulty because there is only counting, no weighting, no "scaling" or measurement problem as exists in the characterization of individuals according to various attributes

From a more substantive point of view, these relational structures are appealing to the social scientist because he sees in them a rare chance to characterize with precision social configurations General experience tells us that there are radical variations in group structures, and that these variations have important consequences for the behavior of people within them For example, an early sociologist concerned with the effects of social configurations on individuals, Georg Simmel, said in discussing social isolation, "The question of whether a group favors or even permits such loneliness in its midst is an essential trait of the group structure itself Close and intimate communities often allow no such intercellular vacuums" (Kurt Wolff (ed) [1950], p 119) Simmel's very language suggests the importance of characterizing a group in terms of its structural properties But without mathematical tools, Simmel was unable to do more than speculate upon the effects of these properties

Our facility in this direction is not much improved even now We know very little about ways to compare group structures, except in the grossest fashion (for example, by the number of in-group choices the group members give, or by the number of mutual choices in the group), and as a result we are hardly able to investigate the effects of social configurations Yet the possibilities are evident, they are exciting ones because of the numerous kinds of investigations they facilitate

(*) It is because in Guttman's scale analysis a particular scale position includes only one pattern of response, not diverse patterns classified together that social psychologists concerned with attitude measurement are generally more satisfied with it than they are with a technique like Likert's

Thus because relational models by-pass the usual problems of measurement, and because they allow sociologists to study important structural problems in sociology, and perhaps for other reasons, they have considerable appeal The promises which sound models hold, *their possible value for* future investigations, appear to rest at least in some degree upon the above two reasons, the first emphasizing the clear outlines of the problem and data for mathematical treatment, and the second emphasizing its importance for social science

Below are some specific comments on each of the approaches which came under the general heading of relational models

1. The approach taken by Rapoport and others in constructing random models appears to be a basic step necessary for the general development of relational models These random models are similar to many others in probability theory, except that it is far from obvious in these cases what relational structure is the expected one on the assumption of random outcome of relations If a coin is unbiased, subject only to random influence when tossed, everyone knows that it will turn up heads very nearly 50 per cent of the time The parameter 0 5, which characterizes an unbiased coin, is a very important one for any work which investigates the amount and type of bias of a coin But it is a parameter which is so evident that it is almost taken for granted The expected configurations (assuming random outcome of pair relations) in relational structures, on the other hand, are not so evident Much work needs to be done in investigating the expected configurations and their variances, so that some standard can be set for indicating in which direction and how much the group structure is 'biased " Knowledge of these expected states is just as important for the development of relational models as is the knowledge that an unbiased coin will turn heads about 50 per cent of the time for the investigation of bias in coins, yet there have been only a few steps in this direction

The work by Rapoport and others shows some of the complexity which occurs in working with these problems, yet even this work is largely confined to expected states, with no investigation of variances Rigorous statistical tests for indicating the direction and amount of deviation from randomness are badly needed Without these, it is hardly possible to say that a group has one or another tendency, *e g*, a tendency toward clique formation

2 The extension of these models to incorporate behavior tendencies deviating in one direction or another from random choice is a second step in the direction of developing adequate ways of characterizing these struc-

tures To go beyond the development of random models to their modification, which allows them to measure certain behavior tendencies, seems an important step in characterizing relational structures In one sense, information theory is one such development, for it offers a means of characterizing the degree of 'disorder' or 'uncertainty" or randomness of a set of elements distributed over some space Whether information theory can be widely applied to these structures is an open question, but in any case some means of characterizing their deviation from complete randomness would be an important addition The few examples mentioned in II do little more than point the way for future development, and are crude instruments, but they begin to suggest some of the values of this approach

3 The next body of work, including that by Luce and Perry and others, deals with an almost prior problem to those discussed in 1 and 2, given a matrix of choices or communication links or dominance relations, what can one say about the structures? How many cliques does it have, or how many *n* cliques, for example? What this approach does, then, is to take the raw data, and process it to discover certain properties of the structure This seems quite as important a step as the others discussed, though it is in a somewhat different direction While the second approach discussed above, the 'modified random" approach, takes as given a particular structure and attempts to infer from it the tendencies which influenced the choices or the dominance relations, this approach simply asks what are some of the properties of the structure? The fact that a mathematical operation is needed to determine even the relatively elementary properties of the structure is an indication of just how complex these relational structures can be

4 The approach illustrated by the several dynamic relational models presented in Section 4 is more nearly what is usually meant by 'mathematical models" than are the other types of relational models These models incorporate a process through which structures may change over time, thereby representing a further stage of development than those of types 1, 2, and 3 above They begin where types 1 or 2 leave off, and attempt to describe processes over a period of time in a group, considering the group as a dynamic system which generates its own equilibrium states The advantage which these models hold over their static counterparts is evident, and will not be elaborated here Instead, I shall make a few comments about the general approach to model-building which these models represent This is best done in the context of a discussion including the other approaches as well But first, the group action models require some comment

Group Action Models The Hays Bush stochastic model of group learning requires comments of several kinds First of all, it is rather unique among these models in that two alternative hypotheses are set down, giving rise to alternative models (*i e*, the 'group actor" model and the 'voting model) which are then *confronted* with experimental data for choosing between them The fact that the experiments do not allow a clear-cut choice should not obscure the nature and value of this approach itself The approach of setting down two (or more) *alternative* models which make explicit the different consequences of different hypotheses about behavior, is useful in itself, quite without experimental confirmation about how groups actually do behave One hypothesis is based on certain assumptions about interaction and influence in a group, a second is based on other assumptions Because each of these assumptions is sometimes met in practice, it is useful to know what their consequences are for the group's behavior

To test between such alternative hypotheses by testing between their consequences is an approach which seems considerably better than the usual one of simply testing the fit of a model If the measure of fit were not intrinsically built into this model, then it would be even more useful to have the two alternatives In such a case the alternatives would allow a better judgment than would otherwise exist of just how *closely either alter native fits the model*

On the other side of the fence, some serious questions can be raised about certain aspects of this model, that is, one might ask just what is learned even if it is found that the model *does* precisely fit the data? Apparently, such a result would tell that the group decision is arrived at through one kind of group decision function (*e g*, the majority rule voting model) But would this really have any usefulness? It is obvious that under various conditions groups reach decisions in various ways, so that such a result could not be given the status of "the decision function by which groups *generally take action* The fact that such a decision function is not some *fundamental and immutable* property of groups is further emphasized by *results found* by Hays and Bush in their experiments the experimental groups *consciously planned* ahead of time how they would arrive at a decision Thus it seems *unrealistic even* to hope that through such experiments one might get at some *underlying*, *fundamental* process of group decision-making

Finally, even taking such a *goal as a realistic* one, why overlay the basic question (how groups come to *reach a decision*) with a complex stochastic

model of learning? Would it not be more productive to set up a simple test situation which would clearly illuminate the decision making process? It seems, in other words, that the overlay of the learning model unnecessarily complicates the basic question which is under investigation

The Lorge Solomon problem solving models constitute an approach somewhat different from any of the others examined here Lorge and Solomon ask themselves, in effect, 'What are the simplest and most reasonable assumptions which will account for the differences between group performances and individual ones?' The models which they set up in answer to this question show clearly the logical consequences of certain very simple assumptions about the way people relate to one another The fruitfulness of such an approach seems evident if data on group performance and individual performance are compared, then these models serve as reasonable standards by which comparisons can be made For example, these models indicate that under the simplest hypothesis of a single-stage problem with no 'group effect," the proportion of groups which solve the problem should be higher by a certain specifiable amount than the proportion of lone individuals who solve it Thus perhaps the major usefulness of these models is much like that of the "random choice' relational models they indicate what results one could expect when there is "no effect" of certain socio-psychological factors They give a basis of comparison or a base line for measurement.

Strategies in Model-building Having discussed each of the different approaches separately, it is useful to examine generally several strategies of model-building, and see just which strategies are represented by the various models

Mathematical models, as all formal theories, have certain attributes first, they are constructs built as analogies or "models" of some phenomenon in the real world The model can be said to consist of four parts, essentially

a) The variables or concepts of the model (including both primitive terms and defined terms),(*)

b) The postulates, which relate these variables in some fashion,

c) The mathematical operations performed on the postulates to obtain deductions,

d) The deductions or theorems which derive from the postulates These

(*) Primitive terms are those not defined within the theory but only outside, in relating it to the real world. Defined terms are simply logical derivatives of primitive terms

are essentially relations between the variables other than those postulated, but nevertheless logically implied by the postulates (*)

These four attributes represent the formal system of the theory In order for the theory to be a theory about real phenomena, it must be so constructed that its structure reflects the structure of relations between actual events At certain points there is a correspondence established between parts of the theory — *i e*, certain of its variables — and the real world This is variously known as measurement, identification, or interpretation

Different model builders operate in quite different ways in their handling of these parts of the model and the correspondence to real phenomena The one difference in operation which seems most important in assessing the potential fruitfulness of models concerns the points at which correspondence is made between the model and the real world There are two extreme methods of making this correspondence

1) The postulates are verified on the basis of observation or experiment, that is, there is correspondence between the model and the real world at the level of the postulates The model is then used to predict phenomena which are *not* yet known or which are thought to be unrelated to (that is, not derivable from) those which served to verify the postulates Theories like that of Malthus (who postulated a simple mechanism of birth frequency, which general experience told him was largely justified, and on the basis of this, predicted population growth curves) are of this sort If we consider not only purely descriptive theories, but normative ones as well, the theory of games is an excellent example of this kind of theory It sets up postulates corresponding as closely as possible to what we feel rational behavior under risk should be, and then deduces the outcome of games in which two or more such rational players oppose one another

2) Certain phenomena in the real world correspond to theorems of a model, though the postulates are unverified Thus the correspondence is on the level of the theorems, and the underlying structure of events is hypothesized to correspond to the model's postulates Such theories as the atomic theory of matter, with atoms and electrons acting only as hypothesized

(*) There is no sharp line between theorems and postulates nor between primitive terms and defined terms for a theory may begin with one set of postulates and prove an other proposition as a theorem, a second formulation may use this last proposition as one of the postulates leaving one of the first set to be proved as a theorem Similary primitive terms and defined terms are often interchangeable Nevertheless a psychological difference often occurs though there is no logical one It may be quite difficult or unrealistic to think of a model as consisting of one set of postulates perfectly natural to think of it as consisting of an alternative set

constructs having consequences consistent with the observable physical world, are of this sort

Where do Simon's model, the models of relative participation rates, the dynamic relational models, and the models of group action stand with respect to this distinction? It is important to locate them in this way, for this gives considerable insight into just what these models propose to do

Simon's model-building is clearly of the first type, the postulates are confirmed and the model's aim is to derive a number of other phenomena from these postulates The goal in Simon's model and others like it is presumably to show that a number of phenomena concerning behavior of groups, which were previously not recognized as related, can be derived from knowledge about only a few factors Thus by making a limited number of observations it is possible, given the theory, to predict that certain other events will occur

The participation-rate models are of precisely the opposite sort The investigators are presented with a regularity produced by some underlying process occurring in the group Thus the task of a model here is to postulate the kind of process which will have as a consequence the same regularity as is actually found The ' statistical model which gave a partial explanation of the regularity was an attempt to mirror the structure of actual events and thus give the same regularity as Stephan found But of course an adequate model would be a dynamic one, which shows how the events occur through time In any case, it is the consequences, not the postulates, to which a correspondence is made Thus if postulates can be found which give as a theorem or consequence this same regularity, it can be hypothesized that these postulates correspond to the actual process Ordinarily, however, such postulates must meet a second criterion as well they must be "reasonable" or "understandable explanations not inconsistent with all one's intuitions

The dynamic relational models are not quite like either of these extreme types of theories Rapoport's and Landau's status hierarchy models start from the observation that (a) there are "peck orders' among chickens, in which a peck relation exists between each pair of chickens, (b) these peck-orders are relatively stable, but sometimes change, (c) small flocks of about ten hens established peck-orders which were near to perfect hierarchies

Thus Rapoport and Landau did not have precise knowledge about the group structure, that is, the consequences of meetings between hens, but they did have the above qualitative knowledge In this sense, their model construction was like that of type 2, for they each experimented with a number

of sets of postulates, attempting to find a set which would generate a dominance structure agreeing with the qualitative information they had

At the same time, Rapoport and Landau had certain information about the postulates themselves They knew that the structure was established through pairwise peck-relations, and that meetings kept occurring and would sometimes result in reversals, although the peck relation was usually maintained

Thus they knew enough about the underlying process involved to be rather certain that a stochastic process of some sort could serve as a model

As a whole, however, the approach of Rapoport and Landau can be considered of type 2, for only certain of the postulates were known, their main goal was to find a set of postulates from which they could derive *the empirically found result The approach is thus not too far different from the* participation rate models of Stephan and others, except that Rapoport and Landau were somewhat less concerned that the consequences of these postulates fit the empirical data This was probably in part because the data they had *were not precise, nor abundant, nor systematically gathered Insofar as* they were not concerned with fitting empirical regularities, nor with confirmation of the postulates, Rapoport's and Landau's approach loses touch with reality The models become "playthings" whose value is hard to determine But this tendency is discussed in more detail below

Leeman's *dynamic model of sociometric choice is not of type 2, because* Leeman begins with postulates, not with a result to be 'explained " *On the* other hand, Leeman has no empirical regularity on the level of the postulates either He starts out neither with a regularity on the level of the model's *consequences, as do the participation rate people* and, to a lesser degree, Rapoport and Landau, nor *with regularities from which he* builds a set of postulates, as does Simon In this respect Leeman's model is like numerous others in social science which are "models" only in a very special sense, for *they are not models of anything* which exists in reality (*)

Most of *Rashevsky's [1951] models* are good examples of this, for they hardly begin to correspond to any *real behavior, they are confirmed at* neither the postulate level nor the consequence *level Suggestive and inter-*

(*) I certainly do *not want to single* out Leeman s work for criticism here as being of less value than *that of other model* builders It is simply that his model exemplifies an *approach which I feel to be relatively* unfruitful For other examples of this same approach *see my paper in* Lazarsfeld [1954 (p 155-165)] in which I construct two such models These models were interesting to construct it is this interest in fact which leads one to forget to ask just how — at best — they might be useful

esting, like mechanical toys, they can find little application to social science as they stand Their main value probably lies in the practice in model-building they give the social scientist or mathematician, and in indicating to him what kind of mathematics may be most amenable to social science problems

Yet, though Leeman's model is developed without any empirical reference, it does not remain so He creates an experimental situation which fulfills some of the postulates, and then compares the deductions of the model with the outcome of the experiment This at first seems to make the model much more empirically valuable, for it is tied down at particular points to phenomena in the real world

However, closer examination gives one little reason to feel that this makes the model more a model of reality, for the "reality' which it attempts to mirror is itself a completely constructed and artificial situation Thus a model is not being constructed to correspond to certain phenomena in the real world, but, conversely, phenomena are constructed to correspond to the model In such a situation, it is difficult to know where the ingenuity lies — in construction of the model, or in manipulation of people so that they will behave in accordance with it It is hard to see just what the goal is in such a model Is it to confirm some postulate about choice behavior, as appears to be the case in Leeman's experiment? If so, construction of the model is superfluous, for the postulate could be better tested directly — and it seems a rather trivial postulate anyway, having to do not with relations in a group, but with the determinants of choice between two objects when there is no real reason for choosing one rather than another It would seem that models having to do with group structures should do more than test a postulate about the psychology of choice behavior It would be one thing to *use* a postulate about choice behavior which had been already confirmed, and to show the consequences of this for group structure, but to attempt to *validate* a postulate about choice behavior through constructing a model having to do with group structure seems a rather roundabout procedure

Leeman's model appears to exemplify well the dangers inherent in model construction followed by experiment It may be that this approach will be ultimately one of great value But until model builders and experimenters are more able to single out just what it is they are testing, and what they will have if they do succeed, they may be better off dealing with empirical regularities which they are forced to treat as given, whether these

regularities derive from experiment or field observation (*) If they succeed in incorporating such regularities in a model, either as postulates or as consequences, they will at least have constructed a model which is tied to phenomena in the real world

The two group action models, the stochastic model of Hays and Bush, and the problem solving model of Lorge and Solomon are obviously of type 2, in which postulate systems are hypothesized to explain or account for certain observed consequences Hays and Bush set up a simple stimulus response situation for the group, observe the group's behavior and then attempt to account for it with one of two alternative sets of postulates The *postulates constitute the hypothesized 'underlying mechanisms of individual and group decision making, and it was these proposed mechanisms which were tested by the experiments*

The Lorge Solomon models are similar, for they took experimental results and posed alternative problem solving processes or the postulates of models designed to account for these results Their work is probably more unambiguously of this approach than any other, for they began with results *of previous experimenters, and thus there was no possibility of their forcing* experimental conditions so as to fit their model, it was clearly a model built to account parsimoniously and reasonably for results which were otherwise not so easily explained

(*) Experiments which are carried out either for their intrinsic substantive interest (*e g*, Festinger's Back's Bales) or prior to construction *of a model to discover* the effect of certain group *structures* (*e g* all the experiments carried out with Bavelas communication structures) certainly are free from this difficulty Such experiments have value apart from the model and there appears to be no danger *that the experimenter is forcing* conditions to fit the *model*

BIBLIOGRAPHY

Asch, S E "Effects of Group Pressure upon the Modification and Distortion of Judgments," in *Groups, Leadership, and Men*, Harold Guetzkow (ed) (Pittsburgh Carnegie Press, 1951), p 170-190

Bales, R F "A Set of Categories for the Analysis of Small Group Interaction," *Am Soc Rev*, *15*, 1950, p 257—263

— *Interaction Process Analysis* (Cambridge Addison Wesley Press, 1951 a)

— "Reply to Keller's Comment," *Am Soc Rev*, *16*, 1951 b, p 843

—, and Borgatta, Edgar F "Interaction of Individuals in Reconstituted Groups," *Sociometry*, *16*, 1953, p 302—320

—, Strodtbeck, Fred L , Mills, Theodore M , and Roseborough, Mary E "Channels of Communication in Small Groups," *Am Soc Rev*, *16*, 1951 c, p 461—468

Bavelas, Alex "A Mathematical Model for Group Structures," *Applied Anthropology*, *7*, 1948 p 16-30

Beum, Jr , Corlin B , and Brundage, Everett G "A Method for Analyzing the Sociomatrix," *Sociometry*, *13*, 1950, p 141—153

Bjerstedt, Åke "Some Examples of the Possibility of Using Structural Formalizations in Sociometric Analysis," *Sociometry*, *17*, 1954, p 68—76

Bush, Robert R., and Mosteller, Frederick *Stochastic Models for Learning* (New York John Wiley and Sons, 1955)

Carter, L. F , and Nixon, M · "An Investigation of the Relationship Between Four Criteria of Leadership Ability for Three Different Tasks," *J Psychol*, *27*, 1949, p 245—261

—, *et al* "A Note on a New Technique of Interaction Recording," *J of Ab and Soc Psych*, *46*, 1951, p 258—260

Cervin, Vladimir "Relationship of Ascendant-Submissive Behavior in Dyadic Groups of Human Subjects to their Emotional Responsiveness," *J of Ab and Soc Psych*, *54*, 1957, p 241-249

Chapple, Eliot D "The Interaction Chronograph· Its Evolution and Present Application," *Personnel*, *25*, 1949, p 295—307

—, and Coon, C S *Principles of Anthropology* (New York Henry Holt, 1942)

Cohen, B P "A Brobability Model for Conformity," *Sociometry*, *21*, 1958, p 69-81

Coleman, James "Independence of Acts in a Social Unit," Technical Report No 2, Behavioral Models Project, Bureau of Applied Social Research, Columbia University, 1953

—, *Applications of Mathematics in Sociology*, to be published by the Free Press

—, and Guttenberg, Helmut "A Machine Method for Obtaining Cycles and Chains in Social Networks,' unpublished, Bureau of Applied Social Research, Columbia University, 1956

—, Menzel, H , and Katz, E „The Diffusion of an Innovation among Physicians," *Sociometry*, *20*, 1957, p 253

Conant, James B ' Robert Boyle's Experiments in Pneumatics," *Harvard Case Histories in Experimental Science*, No 1 (Cambridge Harvard University Press, 1950)

— *On Understanding Science* (New York New American Library, 1951)

Cooley, Charles Horton *Social Organization* (New York Scribners, 1925)
Dahlberg Gunnar *Mathematical Methods for Population Genetics* (Basle, Switzerland S Karger, 1947)
Davis, Robert L. "The Number of Structures of Finite Relations," *Proc Amer Math Soc*, *4*, 1953, p 486—495
— 'Structures of Dominance Relations,' *Bull Math Biophysics*, *16*, 1954, p 131—140
Feller, William *An Introduction to Probability Theory and Its Applications* (New York Wiley, 1950)
Festinger, Leon 'The Analysis of Sociograms Using Matrix Algebra, *Human Relations*, *2*, 1949, p 153—158
— 'Informal Social Communication" *Psychological Review*, *57*, 1950, p 271—292
— 'The Role of Group Belongingness in a Voting Situation,' *Human Relations*, *1*, 1947, p 154—180
Gustafson, Herbert W 'On the Frequency Distribution of Participation in Small Group Discussions," Ph D Thesis, University of Utah 1955
Harary, Frank and Norman, Robert *Graph Theory as a Mathematical Model in Social Science* (Ann Arbor University of Michigan Press 1953)
—, and Ross, Ian "A Procedure for Clique Detection Using the Group Matrix," *Sociometry*, *20*, 1957, p 205—215
Hays, David G 'A Game Learning Model of Interaction, RAND Corporation Santa Monica, California 1956
—, and Bush, Robert R "A Study of Group Action,' *Am Soc Rev*, *19*, 1954, p 693 701
Hemphill, John K "Situational Factors in Leadership ' Ohio State University, *Bureau of Educational Research Monographs*, No 32, 1949 136 p
Henry, William and Guetzkow, Harold 'Group Projection Sketches for the Study of Small Groups," *Journal of Social Psychology*, *33*, 1951, p 77—102
Homans, George C *The Human Group* (New York Harcourt, Brace and Company, 1950)
Horwitz, Murray, and Cartwright, Dorwin "A Projective Method for the Diagnosis of Group Properties,' *Human Relations*, *6* 1953, p 397—410
James, John 'The Distribution of Free Forming Small Group Size," *Am Soc Rev*, *18*, 1953, p 569
Jennings, Helen H *Leadership and Isolation*, revised edition (New York Longmans Green, 1950)
Katz, Leo "The Distribution of the Number of Isolates in a Social Group," *Annals of Math Stat*, *23*, 1952, p 271—276
— "A New Status Index Derived from Sociometric Analysis, *Psychometrika*, *18*, 1953, p 39—43
Keller, Joseph 'Comment on 'Channels of Communication in Small Groups',' *Am Soc Rev*, *16*, 1951, p 842—843
Landahl, H D "A Matrix Calculus for Neural Nets, II' *Bull Math Biophysics*, *9*, 1947, p 99—108
Landau, H G 'On Dominance Relations and the Structure of Animal Societies I Effect of Inherent Characteristics," *Bull Math Biophysics*, *13*, 1951a, p 1—19
— 'On Dominance Relations and the Structure of Animal Societies II Some Effects of Possible Social Factors,' *Bull Math Biophysics*, *13*, 1951b, p 245—262
— 'On Some Problems of Random Nets,' *Bull Math Biophysics*, *14*, 1952, p 203—212
— 'On Dominance Relations and the Structure of Animal Societies III The Condition for a Score Structure," *Bull Math Biophysics*, *15*, 1953, p 143—148
Lazarsfeld, Paul F (ed) *Mathematical Thinking in the Social Sciences* (Glencoe The Free Press, 1954)
Leeman, Calvin P 'Patterns of Sociometric Choice in Small Groups A Mathematical Model and Related Experimentation," *Sociometry*, *15*, 1952, p 220—243

Lewin, Kurt "Forces Behind Food Habits and Methods of Change," *Bulletin of the National Research Council, 108*, 1943, p 35—65

Lippitt, Ronald "An Experimental Study of Authoritarian and Democratic Atmospheres," in *Studies in Topological and Vector Psychology I*, University of Iowa Studies in Child Welfare, 1940, No 16

Lorge, Irving and Solomon, Herbert 'Two Models of Group Behavior in the Solution of Eureka type Problems," *Psychometrika, 20*, 1955, p 139—148

Luce, R. Duncan 'Connectivity and Generalized Cliques in Sociometric Group Structure," *Psychometrika, 15*, 1950, p 169—190

— "A Note on Boolean Matrix Theory,' *Proc Amer Math Soc*, *3*, 1952, p 382—388

—, and Perry, Albert D "A Method of Matrix Analysis of Group Structure," *Psychometrika, 14*, 1949, p 94—116

Masure, R. H, and Allee, W C "The social order in flocks of the common chicken and pigeon," *Auk, 51*, 1934, p 306-327

Minorsky, N *Introduction to Non Linear Mechanics* (Ann Arbor J W Edwards, 1947)

Moore, Omar Khayyam, and Anderson, Scarvin 'Search Behavior in Individual and Group Problem Solving,' *Am Soc Rev*, *19*, 1954, p 702—714

Moreno, J L, and Jennings, H H "Sociometric Measurements of Social Configurations," *Sociometry Monographs*, No 3, 1945, 35 p

Nehnevajsa, Jiri 'Chance Expectancy and Intergroup Choice," *Sociometry, 18*, 1955, p 153—162

Rapoport, Anatol "Cycle Distributions in Random Nets," *Bull Math Biophysics, 10*, 1948, p 145—157

— "Outline of a Probabilistic Approach to Animal Sociology I, II," *Bull Math. Biophysics, 11*, 1949a, b, p 183—196, 273—281

— "Outline of a Probabilistic Approach to Animal Sociology III," *Bull Math. Biophysics, 12*, 1950, p 7—17

— "Nets with Distance Bias," *Bull Math. Biophysics, 13*, 1951, p 85—91

— "Spread of Information Through a Population with Socio-Structural Bias I Assumption of Transitivity," *Bull Math. Biophysics, 15*, 1953a, p 523—533

— "Spread of Information Through a Population with Socio-Structural Bias II Various Models with Partial Transitivity," *Bull Math Biophysics, 15*, 1953b, p 535—546

— "Spread of Information Through a Population with *Socio-Structural* Bias III Suggested Experimental Procedures," *Bull Math Biophysics, 16*, 1954, p 75—81

—, and Rebhun, Lionel I "On the Mathematical Theory of Rumor Spread," *Bull Math. Biophysics, 14*, 1952, p 375—383

Rashevsky, Nicolas *Mathematical Biology of Social Behavior* (Chicago The University of Chicago Press, 1951)

Reich, Edgar "The Game of 'Gossip' Analyzed by the Theory of Information," *Bull Math Biophysics, 13*, 1951, p 313—318

Roethlisberger, F J, and Dickson, W J *Management and the Worker* (Cambridge Harvard University Press, 1939)

Sherif, Muzafir "A Study of Some Social Factors in Perception," *Arch Psychol*, *27*, 1935, No 187

Shimbel, Alphonso "An Analysis of Theoretical Systems of Differentiating Tissue," *Bull Math. Biophysics, 10*, 1948, p 131—143

— "Applications of Matrix Algebra to Communication Nets," *Bull Math. Biophysics, 13*, 1951a, p 165—178

— "A Note on the Problem of Cycles in Random Nets," *Bull Math. Biophysics, 13*, 1951b, p 319—321

— "Communication in a Hierarchical Network," *Bull Math Biophysics, 14,* 1952, p 141—151

Simon, H A "A Formal Theory of Interaction of Social Groups," *Am Soc Rev, 17,* 1952, p 202—211

— *Models of Man* (New York Wiley, 1957)

—, and Guetzkow, H · "A Model of Short- and Long Run Mechanisms Involved in Pressures toward Uniformity in Groups" *Psychological Review, 62,* 1955, p 56—58

Solomonoff, Ray, and Rapoport, Anatol "Connectivity of Random Nets," *Bull Math Biophysics, 13,* 1951, p 107—117

— "An Exact Method for the Computation of Connectivity of Random Nets," *Bull Math Biophysics, 14,* 1952, p 153—157

Stephan, Frederick "The Relative Rate of Communication Between Members of Small Groups," *Am Soc Rev, 17,* 1952, p 482—486

—, and Mischler, Elliot G "The Distribution of Participation in Small Groups An Exponential Approximation,' *Am Soc Rev, 17,* 1952, p 598—608

Tagiuri, Renato "Relational Analysis An Extension of Sociometric Method with Emphasis upon Social Perception," *Sociometry, 15,* 1952, p 91—104

Taylor, D W, and Faust, W L 'Twenty Questions Efficiency In Problem Solving as a Function of Size of Group" *Journal of Experimental Psychology, 44,* 1952, p 360—368

Wei, Telson "On Matrices of Neural Nets," *Bull Math Biophysics, 10,* 1948, p 63—67

Weiss, Robert J *Processes of Organization* (Ann Arbor University of Michigan Press, 1956)

Wightman, W P D *The Growth of Scientific Ideas* (New Haven Yale University Press, 1951)

Wolff, Kurt (ed) *The Sociology of Georg Simmel* (Glencoe The Free Press, 1950)

PART TWO

Survey of Bernoullian Utility Theory

By ERNEST W. ADAMS

University of California

Contents

1 INTRODUCTION

Bernoullian Utility theory(*) falls into the general category of theories of individual preferences and decisions. All of these theories deal in one way or another with the behavior of individuals (usually, though not always, human beings) confronted with the problem of choosing among various alternative courses of action on the basis of the results they expect to follow from these actions and their preferences among the results The present theory is specifically concerned with decision situations in which the outcomes of the actions are not known with certainty in advance, but only with certain degrees of probability Such a situation arises, for example, where a person has to decide whether to accept a 'double or-nothing' offer on a debt owed to him Here he cannot know in advance of taking it what the result of accepting the bet will be, since he may either end up with twice what was originally owed him, or with nothing Bernoullian Utility theory provides an hypothesis which relates the individual's preferences among the possible outcomes together with the probabilities with which each outcome will occur to the individual's ultimate decision as to which action to take

Although the present interest in Bernoullian Utility theory and more generally in theories of decisions involving risks is of comparatively recent origin, this field is now the focus of intense interest, and is undergoing rapid evolution Present interest in Bernoullian Utility dates from the publication in 1947 of the *Theory of Games and Economic Behavior*, by John von Neumann and Oskar Morgenstern In this book, which is already a classic, von Neumann and Morgenstern modernized and made logically precise a theory more than two hundred years old, and used it as a foundation on which to erect their theory of games The theory of utility plays a relatively minor role in the *Theory of Games and Economic Behavior*, since its principal function there is to support the theory of games which is the central subject of the

(*) What we have chosen to call "Bernoullian Utility theory" has no standard name in the literature of decision theory Ward Edwards [1954a] refers to it as the "expected utility theory" and the "expected utility maximization theory" Clyde Coombs calls it the "von Neumann Morgenstern utility theory" [1954] L J Savage [1954] uses the terms "Bernoullian Utility," and "Bernoullian Utility hypothesis" and R D Luce calls it the "linear utility theory" [1957]

book. However, it was immediately recognized that utility theory in von Neumann's and Morgenstern's elegant formulation was an important theory in its own right. Since 1947 Bernoullian Utility theory has had an eventful career, one which has been largely independent of the game theory for which it was originally created. Our object will be to sketch the main developments in Bernoullian Utility theory and its applications since 1947.

The general plan of the survey is as follows. In Section 2 we give an intuitive discussion of the general area of utility theory, including other theories besides the Bernoullian, and attempt to indicate roughly the relationship among the various theories within this area. In Sections 3 and 4 we describe the theoretical framework of Bernoullian Utility, starting with the fundamental postulates, and ending with the derivation of some directly testable consequences which follow from them. Some extensions and modifications of the basic or "classical" Bernoullian Utility model are briefly described in Section 5. Applications of Bernoullian Utility theory are discussed in Sections 6 and 7. These applications are of two kinds: one is as an underpinning of the theory of games, and certain aspects of statistical decision theory (Section 6), and the other is to the explanation and prediction of choice behavior in experimental situations.

The theory of Bernoullian Utility must inevitably be formulated in mathematical terms, and therefore this survey contains a fairly high concentration of mathematical statements. Throughout, however, an effort has been made to keep the mathematical aspects of the theory to a minimum consonant with an adequate presentation of the theory. In particular, the reader should be familiar with such notions as class membership, the elementary set operations, functions, and relations, and the standard notations for these concepts.

The over-all objective of this survey is to present a relatively non-technical introduction to the field of Bernoullian Utility theory which will serve both to acquaint the social scientist with a new and growing discipline which touches on more than one of the traditional fields of the social sciences, and to acquaint mathematically trained non-social scientists with an important application of mathematics in an area which has traditionally been somewhat bare of such applications. In our exposition of the concepts and assumptions of Bernoullian Utility and its applications we shall attempt not only to explain them but also to criticize them, believing that a critical exposition may lead the reader to a deeper insight into these assumptions and the problems which they attempt to solve.

As a survey, this paper does not contain the results of original investiga-

tions Much of the material has, in fact, been dealt with in other surveys Edwards [1954b] has given a brief but thorough account of applications of general utility theory, including our theory, to decision problems in economics and psychology Savage has given a detailed exposition and critique of the assumptions in the various formulations of utility theory and its application to statistical decision theory and the theory of games in *The Foundations of Statistics* [1954], his discussion in many ways paralleling that of Sections 3, 4, and 6 Other papers too numerous to mention present brief, over-all views of the entire field in its relation to neighboring fields in the social sciences However, we hope that the present survey will serve a useful purpose in drawing together the many aspects of Bernoullian Utility theory in a unified and relatively non technical presentation, so that the reader may without too much difficulty gain an integrated and fairly detailed picture of the present status and important trends within this field

2 UTILITY AND ITS MEASUREMENT THE PROBLEM OF SCALING SUBJECTIVE PREFERENCES

2.1 The Basic Concepts of Utility

The fundamental concepts of utility theory, which includes Bernoullian Utility, can best be introduced by way of an example Consider the situation of a buyer in a market faced with the problem of deciding which of a number of possible items to buy For definiteness, we can let our buyer be a housewife, and the market be a grocery store where she is buying groceries Among the things she may buy are such items as meat, vegetables, bread, milk products, etc, and she can buy these in varying quantities and combinations We can think of her in the process of making her decision as mentally reviewing the different possible combinations which she might buy, and weighing these against each other in order to determine which is the 'best The alternatives which she can choose among are limited in two ways, first by the stock actually carried by the grocery store and second by the amount of money the housewife has with which to pay for her purchases Given the amount of money she has and the stock of the store, however, the alternative purchases are perfectly definite, and it remains for her to assess their relative merits in order to decide which one to take This assessment will generally depend on two factors In evaluating a particular combination of items the housewife must take into account the 'intrinsic worth of them to her or her

family and also the price they cost or the amount of money she must give up in order to obtain them. However she does this, she must somehow arrive at a composite estimate or evaluation of each of the possible alternative combinations of groceries which will enable her to rank them in order of preference, and then she must select that one which is highest in this order.

The above account of a buyer's mental processes in the act of deciding which of a number of possible purchases to make brings out some of the essential concepts of decision theory. A typical decision situation involves three things: (1) an individual faced with the necessity of making a choice; (2) a certain set of alternatives among which the individual must choose; and (3) the system of subjective preferences or values by which the individual ranks the alternatives, choosing that one which stands highest according to his values. These three concepts are probably the most fundamental in the theory of decision-making and utility, since they represent elements which are present in almost all kinds of decision situations.(*)

Needless to say, the three elements of decision situations brought out in the preceding paragraph can vary greatly from situation to situation. First, the individual may be almost any sort of thing, which can be interpreted as making choices. For example, he may not be a person at all, but a social or economic institution faced with some sort of policy decision, or he may be an animal such as a rat in a maze. The alternatives too may be of all kinds: possible "baskets" of groceries in a market, possible amounts which a nation can choose to spend on armaments, or different turnings which a rat can make in a maze. Even what appear to be the same sets of alternatives turn out to be different when interpreted in different ways. The housewife's alternatives may either be thought of as possible combinations of items from the grocery store, or as the "acts" of obtaining these items, where the act in-

(*) Some writers have argued, with much justice, that our picture of an individual following a step-wise process of first mentally ranging all the alternatives before him, then assessing their merits, then selecting the most preferred one, then acting, is false, since actual mental processes in most decisions are not carried on with any such quasi-logical formality. In reply to this criticism two points may be made. First, it is almost imperative at the outset in developing any theory of complex phenomena to simplify the picture to the point where the components of it become manageable. Second, from the point of view of behavioristic psychology, it is irrelevant whether or not unobservable subjective behavior such as "decisions" really takes place as we have supposed, since any theory is to be tested on its observable consequences only. It will be seen that even though the concepts of decision and preference denote seemingly unobservable acts and states, theories built on these concepts do have observable consequences. From the behavioristic point of view, the analysis of the mental processes involved in decision simply functions as a heuristic guide to the construction of theories whose scientific meanings lie entirely in their observable consequences.

cludes both the paying over the money to buy the groceries and the receiving of them. Finally, as is obvious, possible preference patterns can vary greatly. Two different housewives with the same amount of money in the same store may end up buying entirely different things. Even the mechanism of preference formation may be radically different in different situations. The preferences of a single individual are dictated in many instances by physical and emotional needs to which preferences are related in very complicated and little-understood ways, while the "preferences" or choices of a social group may be determined by an artificial formal mechanism, such as a voting procedure.

It should be noted too that the three concepts so far isolated do not appear to apply to some situations which are ordinarily regarded as decision situations. Where an individual is engaged in deciding on what are usually called "objective matters of fact" it appears that his personal preferences are not involved. If a person is presented with two weights and asked to pick the heavier of them, it would seem that his choice does not depend on his preferences. Even here, however, it can be argued that the person actually picks the weight in what he hopes is the way which will have the most beneficial results for him. Pressing this argument to its logical conclusion, however, leads to an extreme form of psychological relativism which it is difficult to maintain (at least it would appear to invalidate the very theories we are proposing if they were to imply that their authors simply propounded them as the ones which they expected would have the most beneficial consequences for themselves).

One final concept of utility theory is still to be introduced: that of "utility" itself. The utility of an alternative may be roughly characterized as a measure of the strength of an individual's preference for it. In the case of the housewife, the utility of a particular collection of groceries for her is a numerical measure of its desirability to her. Clearly, the concepts of utility and preference are closely connected, in that the difference in the utility of two alternatives determines which one is preferred to the other. It is usual to speak of an indidual's utility function, denoted "u," where $u(x)$ is a numerical measure of the value or utility of alternative x to the individual. The utility of an alternative can be thought of by analogy to the weight of an object: just as the weight of an object gives a numerical measure of the strength of the earth's gravitational pull on it, $u(x)$ can be thought of as a numerical measure of the strength of the individual's desire for the alternative x. The fundamental assumption relating utility and preference can be stated as follows: if x and y are two alternatives, then the individual

whose utility function is u prefers x to y if and only if $u(x)$ is greater than $u(y)$ This assumption is one of the central assumptions of the theory of utility, including Bernoullian Utility We shall refer to it as the "*ordinal assumption*"

The ordinal assumption by itself does not tell us very much about an individual's preferences and the decision behavior determined by them (we shall see, however, in Section 4 that the ordinal assumption does have some testable consequences) If we suppose ourselves in the position of the external observer of human behavior who can only witness a subject's overt acts, such as his actual choices, then the strengths of the subject's preferences are not accessible to direct observation and can only be inferred from the choices he is actually observed to make It is possible to infer that the subject prefers one alternative to another by observing that he actually chooses it when forced to choose between the two, and this observation in turn allows us to infer that the utility of the chosen alternative is higher than that of the other But how much higher? If the only thing we know about utility is that it is related to preference as postulated in the ordinal assumption, and all we can observe directly is preference, then it is easy to show that there is in fact no objective way for determining utility magnitudes, as opposed to utility differences For example, suppose an individual is observed to prefer alternative x to alternative y, and alternative y to alternative z This information tells us that

$$u(x) > u(y) > u(z)$$

but does not tell us whether the difference between the utility of x and y is greater than, equal to, or less than the difference between the utilities of y and z. Either of the following hypotheses about the utility values of the three alternatives is consistent with observed facts $u(x) = 3$, $u(y) = 2$, $u(z) = 1$, or $u(x) = 1{,}000{,}000$, $u(y) = 1{,}000$, and $u(z) = 1$, and there is apparently no objective way of deciding which of these is correct.

Until the recent revival of Bernoullian Utility theory, it was commonly held that the only "law" of utility theory is simply the ordinal assumption This view, coupled with the opinion that the only observable "facts" are preferences, as revealed by actual choices, leads to the conclusion that statements about utility magnitudes are objectively or empirically meaningless, since there is no way of verifying them This is the view of the "ordinalist school" of utility theorists which largely came to dominate utility theory before the rise of the Bernoullian theory They held that utility is an ordinal concept, since only statements about utility orders (i e, statements as to

whether the utility of one alternative is greater than that of another or not) were thought to be meaningful

If one grants the ordinalist's thesis, that the only law of utility theory is the fundamental ordinal assumption, then it does follow that utility is an ordinal concept, and only utility orderings are objectively significant Bernoullian Utility theory, however, introduces further hypotheses about the relations between utilities, preferences, and alternatives which can be thought of as yielding more information about utility values, making it possible to determine more about them than simply their ordinal relations The rise of Bernoullian Utility theory has also revived interest in other hypotheses about utilities, and each of these hypotheses, or theories, leads to a different way of measuring, or scaling, utilities In the following two sections we shall discuss the general problem of scaling utilities and outline briefly the fundamental ideas that underlie four of the modern approaches to the problem

2.2 Methods of Determining Utility Measures: The Scaling Problem

We have just seen that the ordinal assumption by itself implies that only the ordering of utility values can be inferred from observable phenomena Another way of putting this is to say that according to the ordinal hypothesis utility is only determined as an ordinal scale "(*) In order to obtain a more complete measure of utility, it is therefore necessary to find further observable consequences which follow from a subject's utility values beyond simply the prediction of the ordering of his preferences In this section we shall briefly outline four such hypotheses about utility, each of which leads to a more unique determination of utility values than does the ordinal hypothesis by itself

2.2.1 Utility Difference as a Function of Difficulty of Choice.

The first hypothesis about utility which leads to a stronger scale of utility measurement is the assumption that the subjective "difficulty" of making a decision between two alternatives depends on the difference in utility between them The theory is that the closer two alternatives are together in utility, the greater will be the difficulty experienced by the subject in choosing between them This difficulty may be measured either by the subject's verbal report, or else by observing the length of time it takes him to make his choices This hypothesis was actually used by Coombs

(*) The terms 'ordinal scale,' and linear scale which will occur later, are from the theory of measurement We shall not assume any familiarity with this theory in this report, and shall not be concerned with its concepts The interested reader can find non technical summaries of some of the basic elements of this theory in Guilford [1954] and Coombs, Raiffa, and Thrall [1954]

and Beardslee [1954] in determining a measure of the marginal utility of money. It is clear, of course, that this hypothesis could not be maintained in its full generality, since many factors may enter into a decision situation which might make it extremely difficult for a subject to make a choice. For example, it may be very difficult for a person to make a choice between two jobs, where the utility differences might be very great, and at the same time he would have no difficulty in making up his mind that he would prefer $1.10 to $1.00, even though the difference in utility between the last two alternatives is probably very small. If, however, the alternatives being compared are all roughly comparable in their complexity, this hypothesis is not unreasonable, and has proved a useful method for determining a measure of utility.

That the observation of the subject's difficulty in choosing between two alternatives does lead to a more unique determination of his subjective utility values than does the simple observation of his preferences between pairs of alternatives is easily seen. Suppose, for example, that the subject prefers alternative x to alternative y, and y to z. It then follows that the utility of x is greater than that of y, which is in turn greater than that of z. If it is further determined that the subject's difficulty in choosing x over y is greater than the difficulty in choosing y over z, this implies (if the hypothesis about the relation between difficulty of choice and utility differences is correct) that the utility difference between x and y is less than that between y and z. This information about the relative magnitudes of utility *differences* is essentially new, because it is impossible to deduce it simply from knowledge of the subject's preference ordering.

The experimental procedure followed by Coombs and Beardslee involved observing only direct preferences among alternatives, and relative difficulty of choices. Thus, in addition to observing which of two alternatives a subject preferred among given pairs of alternatives, he determined for certain pairs of choices A and B whether choice A was harder or less hard than choice B. These observations lead to two kinds of information about utility values: ordering of utility values, inferred from choices, and ordering of differences of utility values inferred from the relative difficulties of choices. This kind of information, which is not in general sufficient to give a unique determination of utility values, leads to what Coombs [1951] has termed an "ordered metric" scale of utility. It is perhaps intuitively clear that simply the knowledge of the orderings of utility values and utility differences would not be sufficient as a rule to give a complete determination of utility, for the same reason that one would not expect this kind of information about, say,

comparative length measurements and length differences to yield as much information about length as would be given from measurement with a calibrated measuring rod What is surprising is that, under certain special circumstances, this information is *sufficient* to give a complete determination of utility values

2 2 2 Utility Difference as a Function of Inconsistency of Choice. Another hypothesis about the relation between utility magnitudes and observed choices, very similar to the one just described, is the following In many experiments on subjective preferences, where the subject is required to repeat his decisions between certain pairs of alternatives on many occasions, it is observed that the subject will exhibit a large amount of 'inconsistency' in his choices between certain pairs Thus, for certain pairs, the subject will not consistently choose one over the other, but will choose one a certain percentage of the time and the other the remainder of the time It is then postulated that the degree of inconsistency shown by a subject in several choices between a single pair of alternatives is an index of their nearness in utility value If the subject chooses x over y 60 per cent of the time and y over x 40 per cent of the time, and chooses y over z 80 per cent of the time he is presented with the choice between y and z, while choosing z only 20 per cent of the time on these occasions this may be interpreted as showing that he prefers x to y and y to z, but that the difference in utility between x and y is less than that between y and z Mosteller and Nogee [1951] made an assumption similar to this in analyzing the data they obtained in their experiment on subjects' behavior in gambling situations

There are a number of ways of obtaining a utility scale from statistical data of the kind described above With certain assumptions, these data can actually be used to obtain a unique determination of the utility values of the alternatives Even without additional statistical assumptions, however, these data provide information about the ordering of utility differences, similar to that given by observing the difficulty of choices Hence these data lead to at least an ordered metric scale of utility

2.2.3 Additive Utilities. A third assumption about the relation between utility magnitudes and observed behavior is that when certain alternatives are combined so as to form a "composite' alternative, the utility of the resulting alternative is simply the sum of the utilities of the "component" alternatives out of which it was formed Thus, if the alternatives being considered are various items

and combinations of items in a grocery store, it may be assumed that the utility, say, of a pound of butter and a quart of milk taken together is equal to the sum of the utilities of these items separately The assumption that the utility of composite alternatives is "additive" was generally accepted by classical economic theorists up to and including Marshall (see, for example, Stigler [1950] for a brief survey of the history of the concept of utility in economic theory) The abandonment of this hypothesis, due to the general recognition of the fact that it is obviously false if applied in full generality, led directly to the rise of the *Ordinalist School* of economic theorists and the belief that the only observable consequences of statements about utility are simple preferences The hypothesis has been revived recently, however, and has been used in a restricted form by Ward Edwards [1954b], and by Robert Fagot and the author (Adams and Fagot [1956]), to obtain measures of utility independent of risk

The assumption that utility values add when composite alternatives are formed can be shown to yield information about utility values which does not follow from the utility orderings alone The observation of preferences among composite alternatives implies relations among sums of utility values comparable to the relations among differences between utility values which can be inferred from the difficulty of the choices As in the case of the ordered metric scales, these relations among the sums are usually not sufficient to give a unique determination of the utility values of the alternatives, though under special circumstances these relations may be sufficient to determine the values uniquely

2.2.4 Expected Utility.

The last hypothesis to be discussed about relationships between subjective utility and observed choices involves the assumption that the utility values of alternatives which are in a sense "certain" are related in a definite way to the utility values of probability combinations, or "mixtures," of these alternatives The intuitive idea underlying this hypothesis is illustrated in the following example Suppose that a man *A* is offered a bet by another man *B* in which *A* wins $10 from *B* if a fair coin falls heads when flipped, and if the coin falls tails, *A* pays *B* $10 This is a decision situation in which *A* has two alternatives to accept the bet, or to reject it If *A* rejects the bet, the outcome is certain he keeps the money he had before the decision, neither winning nor losing If *A* accepts the bet, the outcome is uncertain if the coin falls heads, he wins $10, and if it falls tails he loses $10, but he cannot be sure before making his decision which of these two uncertain out

comes will occur According to the classical "ordinal theory of utility, *A* will choose whichever alternative has the higher utility, and hence must compare the utility of the sure outcome of *not betting and neither winning nor losing*, with the utility of the *uncertain alternative in which* the possible outcomes are winning $10 and *losing $10 It is* reasonable to assume that the utility of *the uncertain alternative* of accepting the bet will depend in some way *on the utilities of* the possible outcomes of choosing this alternative — *i e*, *on the utility of* winning $10 and the utility of losing $10 — as well as *on the relative* probabilities with which these outcomes will occur Whatever relationship is assumed to hold between the utility of the uncertain alternative of accepting the bet and the utilities of its two possible outcomes and their relative probabilities, it is easily seen that the utility of accepting the bet will depend on the utility magnitudes of the possible outcomes, and not solely on their ordering Thus, in determining whether *A* accepts the bet, it is not sufficient to know that he would prefer winning ten dollars to getting nothing, and prefer getting nothing to losing ten dollars, since clearly his choice will depend on *how much more* he prefers gaining ten dollars to getting nothing, and how much more he prefers to hold on to what he has than to lose ten dollars

Although there are several conceivable hypotheses about the relation between the utility of an alternative with uncertain outcomes and the utilities of those outcomes and their probabilities, probably the most intuitively acceptable, and certainly the simplest is that the utility of an *uncertain alternative is the statistical expected value* of the utilities of its possible outcomes *This is the Bernoullian Utility* hypothesis advanced over two centuries ago by Daniel Bernoulli [1954] *and recently revived* and mathematically reformulated by von Neumann and Morgenstern [1947], and with which we shall be concerned *throughout the remainder of* this survey In the situation described above, the Bernoullian *Utility hypothesis* asserts that if the utility to *A* of winning ten *dollars is* $u(10)$, *and the* utility of losing ten dollars is $u(-10)$, *then the utility of the alternative of accepting* the bet and taking a 50 50 chance of winning ten *dollars and losing* ten dollars is $0.5u(10) + 0.5u(-10)$, *i e*, the utility of the gamble *is the expected* value of the utilities of its possible outcomes *If the theory is correct, then A* should accept the bet if $0.5u(10) + 0.5u(-10)$ *is* greater than $u(0)$, and should reject the bet if it is less than $u(0)$

If the Bernoullian Utility hypothesis is correct, then *it is theoretically* possible to determine a unique measure of utility, once *an arbitrary* unit of utility measurement and zero on the utility scale are fixed Suppose, for

example, that the outcome of neither gaining or losing money is selected arbitrarily as having zero utility, and the outcome of losing ten dollars is defined as a loss of one 'utile," or unit of utility measurement It is thus stipulated that $u(0) = 0$, and $u(-10) = -1$ To determine the utility to A of winning ten dollars, it is only necessary to determine the probability, say p, at which A will be undecided as to whether or not he should accept a bet with probability p of winning ten dollars and probability $1-p$ of losing ten dollars If this indecision is interpreted to mean that the alternatives of accepting or rejecting the bet have equal utility to A, then it follows that for this value of p, $pu(10) + (1-p)u(-10) = u(0)$ Since $u(0) = 0$, and $u(-10) = -1$, the above equation implies that $pu(10) + (1-p)u(-10) = 0$, or $u(10) = (1-p)/p$, and hence $u(10)$ is determined

2.3 Risk, Probability, and Uncertainty.

The idea of *risk* is not involved in the concepts of general utility theory which we have been discussing, however, it is of the utmost importance to the theory of Bernoullian Utility Risk or probability enters in decision situations in which the subject is required to choose among alternative courses of action, the outcomes of which he cannot predict with certainty The simplest and most clear cut situations of this kind arise in gambling In a game of poker, for example, a player may be required at some point to decide whether or not to "call" his opponent (for non-poker players, to "call" an opponent means to bet an amount of money equal to what the opponent has already bet) If the man who "calls" has a better hand than the opponent, then he wins, and otherwise he loses, but he cannot know in advance of deciding on the alternative of calling whether his hand is better or not Gambling games furnish obvious examples of alternatives involving risk, but such alternatives are common in decision situations of all kinds The man buying life insurance does so against the risk that he may die before reaching old age, but he doesn't know when his death will occur A military commander must often decide on a strategy or tactic without knowing what the opposing commander's plans are, and hence without knowing certain factors which would enable him to predict the result of his maneuver with certainty A little reflection convinces one that in actuality most decisions must involve an element of risk, even though these risks are often practical certainties Since Bernoullian Utility theory recognizes this risk factor in decisions explicitly, and seeks to specify how the risks involved in an alternative affect its subjective value, it is well to examine this concept

To get closer to the concept of risk, we may begin with an alternative

"action" A, which has several possible consequences, C_1, C_2, C_3, , none of which can be eliminated as impossible in advance of actually taking action A For example, action A might be the act of rolling a single die, the possible consequences would be the various ways in which the die might come to rest — *i e*, there would be six possible outcomes of this action, and the actual outcome cannot be known in advance of rolling the die Now, several cases arise First, it might happen that it is possible to specify definite *probabilities* for the outcomes C_1, C_2, C_3, , without being able to state which one will happen The example of rolling the die falls into this category, since the probability for any given face coming up is 1/6 (at least if the die is "fair") The term '*risk*' is usually reserved to be applied to situations in which the outcomes of the alternatives have definite probabilities A second type of situation occurs where it is not possible to define probabilities for the possible outcomes of A in any clear cut way Suppose that a subject is told to put his hand into a bag and draw forth a ball, which he is told in advance will be either black or white The possible outcomes here are just drawing a white ball and drawing a black ball, but the subject is given no information as to the relative probabilities of the two In this situation the true probabilities are unknown (in some theories it is argued that the "probability' of drawing a white ball given no advance information is 1/2), and hence this would not be regarded as a "risk situation' in the narrow sense of that term Instances in which the subject is unaware of the actual probabilities are called "*uncertainty*" situations

The above dichotomy of alternatives in which the outcomes are not known with certainty into "risks" and "uncertainties must be complicated in two ways In the first place, there may be alternatives in which the subject does not know what the probabilities of the outcomes are, but acts as though he believed that they had certain definite probabilities In such a case it would be said that the person making the decision had a *subjective probability*," even though there were no objective probabilities, or the person was unaware of them Secondly, there are situations in which not only does the person deciding not know the probabilities of the outcomes of his actions, but which are further complicated by the fact that the final outcome will depend on another decision which may be made after the first one, and in which the person making this second decision knows what the first person decided on Instances of this type occur frequently in competition Competition situations are distinguished from ordinary uncertainty situations by the fact that in ordinary uncertainties (such as in the case of the man drawing the ball from the bag), the outcome depends only on the

actual "state" of nature (in the example of drawing the ball from the bag, this state would simply be the actual color of the ball in the bag, which does not depend on the decision of the person), whereas in competition situations the outcome will in general depend not only on the state of things but also on the decisions of another person The decisions of the military commander involve this element of competition, since the final result of his decision depends on other decisions (by the enemy commander) as well

Bernoullian Utility theory as originally formulated was meant to apply only to risk situations, and not to uncertainties Thus, this theory was applied only to choices among alternatives with possible outcomes whose respective probabilities are known to the decider That the restriction to risk alternatives is extremely narrow is probably obvious, because, aside from a few highly artificial situations arising in gambling games, very few important kinds of actions have outcomes whose probabilities can be specified with reasonable exactness Recent opinion on experimental applications of Bernoullian Utility theory has tended to favor applying it to uncertainty situations in which it is reasonable to assume that the subject has subjective probabilities Even in instances in which there is a definite set of objective probabilities for the outcomes of an alternative, it would seem sometimes to be more plausible to interpret the probabilities involved as subjective rather than objective

Readers familiar with the philosophical controversies surrounding the concept of probability may wonder at our cursory exposition of it in view of the fact that it is central to the theory of Bernoullian Utility It can only be replied that Bernoullian Utility theory, like other theories involving statistical or probablistic hypotheses, cannot clear up the problems underlying the definition of probability, but must make do with the vague and ill-defined conceptions current at present It will be noticed further on that in applying the theory experimentally to situations involving *subjective* probability, the problem of defining probability is bypassed, since the determination of real or objective probabilities is not involved In our initial explanation of the assumptions of Bernoullian Utility theory, however, we shall simply assume numerical values for the probabilities of the outcomes of an action to be defined in some way, and proceed from that point without deeper inquiry into the meaning of probability

3 BERNOULLIAN UTILITY THEORY

3.1 The Set of Alternatives: Mixture Spaces

As previously noted, there is a great variety of situations which can be subsumed under the theory of utility What is common to them all is the set of fundamental concepts in terms of which they are described, and what is needed now is a systematic notation for these concepts One element in all decision situations is a set of alternatives among which the decider must choose These alternatives may be of all kinds — from possible bundles of groceries to political officials chosen by an electorate In the general utility theory no special assumption is made about the nature of the alternatives, and the only thing that is presumed is that the possible alternatives are clearly specified and constitute a precisely defined set K For many purposes in utility theory it is not necessary to inquire further into the nature of the members of K, and the assumptions of the theory are stated simply in terms of the members of an abstract set K, called for convenience the '*set of alternatives*'

The theory of Bernoullian Utility is concerned specifically with decisions among alternatives of a special kind namely, alternative actions of which the outcomes can only be known with definite probabilities One such alternative was discussed previously, in which a man has the option of betting $10 on the fall of a coin, to win $10 if it lands heads and to lose $10 if it lands tails This alternative may be thought of as one in which there is a 0 5 probability of getting the outcome $10 and a 0 5 probability of getting the outcome of losing $10

Such an alternative is called a 'probability mixture' of its possible outcomes of winning and losing $10, which are called the 'pure outcomes' In this section we shall describe a systematic notation for probability mixtures, and probability mixtures of other probability mixtures

Suppose that x and y are two possible outcomes or alternatives (it may help to think of them at this point as being pure alternatives, though this is not necessary), and we wish to represent the alternative of taking a chance with probability p of getting x or otherwise getting y If this alternative is chosen by some person A, then A will receive exactly one of x or y, and the probability he will get x is p and the probability he will get y is $1-p$ ($1-p$ being simply the probability he *will not* get x) This alternative can be represented by an ordered pair or "vector," $<px, (1-p)y>$ (*) Through

(*) The notation we use for mixtures is not standard A more common way of representing the alternative of getting x with probability p or else y is xpy We use the present notation because it can be generalized in a more straightforward way to probability mixtures with more than two pure outcomes

out most of what follows it will be assumed that the set of alternatives K satisfies the condition that if it contains two alternatives x and y, then it contains all probability mixtures $<px, (1-p)y>$ (where p is a probability, hence satisfies the condition $0 \leq p \leq 1$)

In interpreting a probability mixture $<px, (1-p)y>$ it must be kept in mind that this represents the alternative of "gambling" with a probability p of getting some outcome x, and probability $(1-p)$ of getting y In particular, the combination px in the symbol $<px, (1-p)y>$ must not be thought of as an arithmetical product of two numbers p and x

Consideration of the interpretation of probability mixtures leads to the postulation of certain laws relating them In general, two probability mixtures $<px, (1-p)y>$ and $<qa, (1-q)b>$ are interpreted as being the same if both yield the same outcomes with the same probabilities An obvious example of two mixtures that yield the same outcomes with the same probabilities are the two alternatives $<px, (1-p)y>$ and $<(1-p)y, px>$ The first of these is the alternative which yields alternative x with probability p and y with probability $(1-p)$, while the second is the alternative which yields y with probability $(1-p)$ and x with probability p Hence, we can immediately assert that for all x and y in the set K, and probabilities p,

$$<px, (1-p)y> \;=\; <(1-p)y, px> \tag{1}$$

Similarly, if we consider the two alternatives $<px, (1-p)x>$, and x, we see that the first yields a probability p of getting x and $(1-p)$ of getting x, hence yields a certainty of getting x Therefore $<px, (1-p)x>$ is regarded as being the same as x

$$<px, (1-p)x> \;=\; x \tag{2}$$

Besides the two rather obvious laws stated above, there are two others of importance Consider the alternative $<px, (1-p) < qy, (1-p)z>>$, which is a probability mixture of x with probability p and the mixture $<qy, (1-q)z>$ with probability $1-p$ The three possible outcomes of this alternative are simply x, y, and z, and the laws of probability indicate that x has probability p, y has probability $(1-p)q$, and z has probability $(1-p)(1-q)$ Now, if it is supposed that not both p and q are equal to zero, then it is easily seen that the alternative

$$<(p+p-pq) < \left(\frac{p}{p+q-pq}\right)x, \left(\frac{q-pq}{p+q-pq}\right)y >, (1-p)(1-q)z >$$

yields the probabilities p for x, $(1-p)q$ for y, and $(1-p)(1-q)$ for z, and hence is the same as the alternative $<px, (1-p) <qy, (1-q)z>>$ There-

fore it can be asserted that for all alternatives x, y, and z in A, and for all probabilities p and q not both zero,

$$< px, (1-p) < qy, (1-q)z>> = \tag{3}$$

$$< (p+q-pq) < \left(\frac{p}{p+q-pq}\right) x, \left(\frac{q-pq}{p+q-pq}\right) y >, (1-p)(1-q)z>$$

Finally, suppose that for some probability $p \neq 0$, the two alternatives $<px, (1-p)y>$ and $<pz, (1-p)y>$ are equal If both of these two alternatives yield the same outcomes with the same probabilities, and there is some chance that x and z can occur (since $p \neq 0$), it can only be that x and z are equal Therefore another law of probability mixtures is that if

$$<px, (1-p)y> = <pz, (1-p)y>, \text{ and } p \neq 0, \text{ then } x = z \tag{4}$$

The four laws of probability mixtures given above which follow from the rule that two probability mixtures of outcomes are equal if each of them gives the same outcomes with the same probabilities may be taken as axioms for systems of alternatives which include probability mixtures of outcomes Such a system of alternatives is an example of what is called a "*mixture space*," the notion of a mixture space being simply a mathematical formalization of the intuitive notion of probability mixtures of alternatives as described above, and satisfying the four laws given A mixture space is formally defined as follows

Definition 1. (*) ***(Mixture Spaces)*** A set A which satisfies the following five axioms is a mixture space

M1 For all x and y in A and probabilities p (*i e*, real numbers p such that $0 \leq p \leq 1$), $<px, (1-p)y>$ is a member of A

M2 For all x and y in K and $0 \leq p \leq 1$,

$$<px, (1-p)y> = <(1-p)y, px>$$

M3 For all x, y, and z in A, and all $0 \leq p \leq 1$ and all $0 \leq q \leq 0$, such that not both p and q are zero,

$$< px, (1-p) < qy, (1-q)z>> =$$

$$< (p+q-pq) < \left(\frac{p}{p+q-pq}\right) x, \left(\frac{q-pq}{p+q-pq}\right) y >, (1-p)(1-p)z>$$

(*) These axioms are due to Melvin Hausner [1954] The content of Hausner's axioms are the same as the ones given here though his symbolism differs slightly from ours

M4 For all x in K and all $0 \leq p \leq 1$,

$$<px, (1-p)x> = x$$

M5 For all x, y, and z in K, and $0 < p \leq 1$, if

$$<px, (1-p)y> = <pz, (1-p)y>,$$

then $x = z$

In the systems of alternatives discussed above each alternative can have at most two outcomes, though each of these outcomes may itself be another mixture with two possible outcomes Thus, the two possible outcomes of the alternative $<px, (1-p)<qy, (1-q)z>>$ are x and $<qy, (1-q)z>$, though the second of these is itself a mixture with possible outcomes y and z For some purposes it is useful to generalize the notion of a mixture to include alternatives with more than two possible outcomes Such an alternative might arise in betting on a horse race, in which the bettor wins a certain amount A in case his horse wins, an amount B, if his horse comes in second, and loses the amount C of his bet if his horse finisheds third or later In this case there are three possible outcomes, A, B, and C, and logically exactly one of these three must occur Assuming that it is possible to assign probabilities to the outcomes A, B, and C, then the risk alternative of making the bet can be represented as a probability mixture $<pA, qB, rC>$, in which p is the probability the horse will win, q is the probability that he will place second, and r is the probability that he will place third or later The only requirement which must be met by the three probabilities p, q, and r is that their sum must be one, since they represent the probabilities of three mutually exclusive events, one of which is bound to happen

It is possible to generalize the notion of a mixture space defined above to include all probability mixtures of two or more alternatives, or even of an infinite number of alternatives The axioms for such mixture spaces would be straightforward generalizations of the axioms given for mixtures with two outcomes Generally speaking, two mixtures would be construed as the same if each one yielded the same possible outcomes with the same probabilities The extension to mixtures of an arbitrary finite number of alternatives actually does not increase the number of possibilities beyond those included in mixtures of just two possible outcomes Thus, the mixture $<pA, qB, rC>$ of three possible outcomes can be represented as the two-outcome alternative

$$<pA, (1-p)<\frac{q}{1-p} B, \frac{r}{1-p} C>>,$$

provided $p \neq 1$, and if $p = 1$, this alternative is equivalent to A We shall not develop the laws of mixtures of arbitrary numbers of possible outcomes here, but defer a more extended discussion to a later section in which we give some axioms for Bernoullian Utility theory as applied to such alternatives

In concluding this section on probability mixtures and mixture spaces, we shall discuss briefly a variation on the mixture concept, in which in place of the probability p in the mixture $<px, (1-p)y>$, we put the event of which p is the probability To illustrate this idea, we can return to the example in which a man is considering betting \$10 on the fall of a coin, to win \$10 if it falls heads, and lose \$10 if it falls tails In the ordinary mixture notation, this alternative might be represented $<5\,(\$10),\ 5(-\$10)>$, since the probability of the coin's falling heads is 5 This notation has what might be thought of as the 'defect of concealing the actual event on which the outcome of the bet depends, and exhibiting only its probability This defect can be repaired by replacing the probabilities p and $1-p$ by the actual events of which p and $1-p$ are the probabilities of occurring In the example, the event on which the outcome of the bet depends is the fall of the coin, which may happen in one of two ways it may fall heads, which we shall denote H, or it may fall tails, which will be denoted 'T Now, a way of denoting the alternative of betting which exhibits explicitly the event determining the outcome is simply to replace the probabilities in the probability mixture by H and T, thus $<H(\$10),\ T(-\$10)>$ In this example, the event T is simply the event of H's not happening, which can be denoted by writing a horizontal bar over H In this notation, the alternative of betting is denoted $<H(\$10),\ \overline{H}(-\$10)>$ In the general case of a probability mixture $<px, (1-p)y>$, in which p is the probability of some event E, and $1-p$ is the probability of the non occurrence of E (denoted $\overline{E}$), the mixture alternative may be denoted $<Ex, \overline{E}y>$

The advantage of this second method of representing mixtures is that it avoids some of the very serious difficulties often inherent in assigning a numerical probability to the occurrence of an event In the case of betting on the fall of a coin, there is a little difficulty in assigning a probability to its falling heads, but there are many events (*e g* the outcome of the 1960 presidential elections) in which there is by no means a clear cut solution to the problem Even where objective probabilities can be assigned, it is often not reasonable to assume that the person making the decision is aware of what they are and acts in accordance with them

A disadvantage of replacing the probabilities p and $1-p$ by the events E and $\overline{E}$ is that in formulating and justifying the assumptions of the theory of

Bernoullian Utility, essential use is made of the laws of probability combination. If the probabilities are replaced by events it is not clear what laws, if any, there are analogous to those of probability for events Nevertheless, there are some decision theories (notably that of L J Savage [1954]), which are founded on this second conception of a probability mixture We shall examine one of these briefly in Section 5

3 2 Preference Relations Among the Alternatives.

Let us now suppose a subject confronted with some set K of alternatives among which he will have certain preferences The subject's preferences are represented mathematically by a "*preference relation*" P in the following sense if the subject prefers an alternative x to another alternative y, we write

$$xPy$$

As far as the subject's preferences among two alternatives x and y are concerned, there are only three possibilities either he prefers x to y, or he prefers y to x, or he has no preference between them In the first two cases, we have either xPy or yPx If the subject neither prefers x to y nor y to x, we will assume that this means that the two alternatives are equally preferable to him, or that he is "indifferent" between them We may represent the case in which the subject is indifferent between x and y by writing

$$xIy$$

The relation I will be called the subject's "*indifference relation*"

For some purposes it is more convenient to define both the preference relation P and the indifference relation I in terms of a single relation R called the "*preference-or-indifference relation*" The relation

$$xRy$$

between two alternatives x and y is defined to hold if and only if the subject either prefers x to y, or feels indifferent between them Both the relations P and I can be expressed or defined in terms of the preference-or-indifference relation R Thus, if xRy holds, then the subject either prefers x to y or is indifferent between them, hence he *does not* prefer y to x Therefore it must be the case that the subject prefers y to x if and only if the relation xRy does not hold This means that the preference relation can be "defined" in terms of the preference-or indifference relation Similarly, if the subject is indifferent between x and y, then both xRy and yRx must hold, and conversely, if both xRy and yRx hold, then the subject must be indifferent between x and y

These observations lead to the following two definitions(*) of P and I in terms of R

Definition of the preference relation For all x and y in K, xPy holds if and only if yRx does not hold

Definition of the indifference relation For all x and y in K, xIy holds if and only if both xRy and yRx hold

3.3 Utility Functions.

The one important concept remaining to be introduced is that of "utility" itself Intuitively, utility is thought of as some sort of numerical measure of the strength of a subject's preference for an alternative Mathematically, utility is represented by a "*utility function*," u If K is the set of alternatives, then u is a function whose domain is K — i e, u has a value corresponding to every alternative x in K — and its values are real numbers If x is any element of K, then $u(x)$ is a real number, called ' the utility of x," which represents the strength of the subject's preference for x Nothing is specified as yet as to how the values of u are to be determined It will be noticed that, unlike the relations P, I, and R, which have fairly clear objective meanings, the objective meaning of asserting "the utility of alternative x is $u(x)$" is extremely unclear We shall defer discussion of the empirical interpretation of the utility function to later sections

3 4 The Basic Assumptions of Bernoullian Utility Theory.

The intuitive assumptions on which Bernoullian Utility theory is based were discussed briefly in Section 2, and these can now be formulated mathematically in terms of the set K, the relations P, I, and R, and the function u The first assumption to be stated is not peculiar to Bernoullian Utility theory, but is common to almost all theories of utility It simply specifies that the utility function must reflect the ordering of the alternatives in preference Thus, one would expect that if alternative x is preferred to alternative y, then the utility of x should be greater than that of y In terms of the preference relations and the function u, this assumption is formulated for all x and y in K, if xPy, then

$$u(x) > u(y) \cdot$$

Similarly, one would expect that if the utilities of two alternatives were

(*) Whether these should be called definitions or axioms may be debated It will depend on the interpretations given to P, I, and R in specific instances whether these definitions really assert equivalences of meaning, or even whether they are true

equal, then they should be indifferent in the subject's estimation for all x and y in K, if

$$u(x) = u(y),$$

then xIy

Somewhat more questionable assumptions are the converses of the two stated above These are that if the utility of one alternative is greater than that of another, then the first alternative should be preferred to the second, and if two alternatives are indifferent, then their utilities should be equal These assumptions are doubtful if the preference relations are to be given straightforward empirical interpretations, since they imply that the subject is perfectly sensitive to utility differences, and should be able to say, for example, that he prefers x to y when $u(x) > u(y)$ no matter how close x and y are to each other in utility value These converse assumptions imply that there is no "just noticeable-difference" level, such that if two alternatives differ in utility by an amount smaller than the j n d , then the subject would declare them to be indifferent Nevertheless, as an approximation, these assumptions may be acceptable

All of the four assumptions about the relation between the utility function and the preference relations can be condensed into a single hypothesis about the relation between u and the preference-or-indifference relation R We shall call this assumption the "ordinal assumption," since it is the principal hypothesis of the theory of ordinal utility

Assumption 1 (ordinal assumption).
For all x and y in K, xRy if and only if

$$u(x) \geq u(y) .$$

That the ordinal assumption implies the four hypotheses stated previously about the relation between the utility function and the relations P and I follows from the definitions given in Section 3 of P and I in terms of R To show that xPy implies that $u(x) > u(y)$, we make use of the fact that xPy holds if and only if yRx does not hold Therefore, if xPy, then yRx does not hold, $u(y)$ is not greater than or equal to $u(x)$, and hence $u(x) > u(y)$ If xIy, then according to the definition of I, both xRy and yRx hold, and hence, according to the ordinal assumption, $u(x) \geq u(y)$ and $u(y) \geq u(x)$ are both true, and therefore $u(x) = u(y)$

The postulate or assumption peculiar to Bernoullian Utility theory is that the utility value of a probability mixture of alternatives is the expected value of the utilities of the alternatives We shall call this the "expected utility assumption "

Assumption 2 (expected utility assumption).

Let K be a mixture space For all x and y in K, and all $0 \leq p \leq 1$,

$$u(<px, (1-p)y>) = pu(x) + (1-p)u(y)$$

This assumption may, of course, be generalized to apply to mixture spaces in which it is possible to form mixtures of arbitrary finite numbers of alternatives If $x_1, \ldots, x_n$ are alternatives, and $p_1, \ldots, p_n$ are probabilities which sum to 1 (i e , $p_i \geq 0$ for all $i = 1, \ldots, n$, and $p_1+p_2+ \ldots +p_n = 1$), then the assumption is that the utility of the mixture $<p_1x_1, \ldots, p_nx_n>$ is

$$u(<p_1x_1, \ldots, p_nx_n>) = p_1u(x_1) + p_2u(x_2) + \ldots + p_nu(x_n)$$

It may not be clear on immediate inspection what the empirical meanings of Assumptions 1 and 2 are, since nothing has so far been said about the interpretation of u The empirical meanings of these assumptions must depend on two things first, a more precise specification of the meanings of the fundamental concepts in terms of which the assumptions are formulated, and second, an analysis of the logical consequences of the assumptions In the following section we shall show that there are a number of possible interpretations for the fundamental concepts, and therefore for the assumptions, each of these interpretations will actually define a different theory, though each theory will have the same formalism It is not, of course, possible to understand the interpretations of the assumptions, without understanding their logical consequences, and the most important of these will be developed in Section 4

3.5 Possible Interpretations of the Concepts and Assumptions of Bernoullian Utility Theory.

Each of the basic elements of the theory of Bernoullian Utility — the set K, the relations P, I, and R, and the function u — can be interpreted in several ways, and hence the number of possible combinations of interpretations is extremely large What we shall do in this section is to indicate some of the major *kinds* of interpretations of the basic concepts and assumptions, without attempting to enumerate all the possible variations Many of these variations, and certain problems of interpretation which cannot be entered on in this section will become apparent in later sections in which we describe some of the applications of Bernoullian Utility theory, both as an adjunct to other theories of behavior, and in experiments designed to test the hypothesis of this theory

3.51 Interpretations of the Set K.

In Bernoullian Utility theory, two problems arise when one attempts to give a precise definition of the elements of the set of alternatives. The first problem is to determine just what is to constitute an "alternative," and the second arises in connection with the interpretation of the probabilities which are involved in the mixture alternatives.

In any particular application of utility theory to a decision situation in which a subject is required to choose among a set of alternatives, it is usually possible to describe the alternatives in more than one way, and this may give rise to ambiguity in their definition. For example, in the situation faced by the man who must decide whether or not to accept a bet in which he wins ten dollars or loses ten dollars, each with a probability of $^1/_2$, it is possible to describe the possible outcomes in several ways. Suppose, for example, that the man has at that time a total of $100 in cash. Is the outcome of winning ten dollars to be described as "gaining ten dollars," or as "having a total of one hundred and ten dollars"? This distinction may seem unimportant, but in fact it becomes critical in many applications. In most experimental applications of utility theory, it must be assumed that the utility of "the same" alternative is the same or nearly constant throughout the period in which the experiment is carried out. But the assumption that the utility to a subject of having $110 is substantially constant during an experiment is entirely different from the assumption that gaining ten dollars will always represent the same gain in utility. In applications, then, in which it must be assumed that the utilities of certain kinds of alternatives remain nearly constant, the identification of certain alternatives as "the same" may pose a problem. In a given experiment, it may be possible to specify *identity* in more than one way, and each such way actually represents a different interpretation of the set K, and hence defines a different meaning for the fundamental assumptions.

Once the notion of identity has been defined, the problem of interpreting the probabilities involved in the mixture alternative arises. One might expect that the dependence of the set K on probability would lead to difficulties in interpretation, since the logical and philosophical question of the meaning of probability is in a very unsatisfactory state (see, for example, Harold Jeffries [1957]). We cannot go into the various interpretations of the logical notion of probability here, but it is clear that each definition of probability leads to a different interpretation of the probability mixtures, and ultimately to a different interpretation of the expected utility assumption.

Beyond the various logical interpretations of probability, the question arises as to whether the probabilities involved in the mixtures should be interpreted as logical probabilities at all It is argued that if Bernoullian Utility theory is to describe actual behavior, then the numbers p and $1-p$ in a mixture alternative $<px, (1-p)y>$ should not represent what might be called "objective probabilities, but should represent what the subject thinks the objective probabilities are *i e*, p and $1-p$ should represent subjective probabilities If p and $1-p$ are interpreted as subjective probabilities, then the difficulties of applying the theory to a particular situation are increased It becomes necessary to determine not only the utility values of the alternatives, which represent subjective magnitudes, but also the subjective probabilities Nevertheless, it is clear that in a situation of any complexity, the subject will be largely unaware of what the objective probabilities of the outcomes of the alternative courses of action are, and if the theory is to be applied at all, some provision must be made for interpreting probability subjectively We shall see that in the experiments designed to test the hypotheses of Bernoullian Utility theory, great effort has been made and many ingenious devices have been introduced to control the subjective factor in the determination of the probabilities in mixture alternatives

3 5 2 Interpretations of the Preference Relations

Of the three basic elements of utility theory — the set A, the preference relations, and the function u — the interpretation of the preference relations offers the least difficulty Nevertheless, there are substantial differences in the interpretations of the preference relations which need to be discussed if a clear understanding of the meaning of the utility assumptions is to be gained The simplest interpretation of the relation P asserts that xPy holds if and only if the subject chooses x in preference to y whenever he is offered a choice between those two alternatives only This interpretation is simple in that it gives a clear cut objective criterion for determining whether or not x is preferred to y Unfortunately, this definition of the relation P is for many applications unsatisfactory First, it may happen that the subject is never presented with a choice between x and y alone, and hence the relation which must hold between x and y would in such a case be indeterminate according to this definition Such a conclusion would in many instances contradict the assumptions of utility theory which allow the possibility that the subject's preference between two alternatives can be inferred from other choices, though he may never have been presented with a choice between them alone For example, one of the consequences of Assumption 1 is that if

xPy and yPz, then xPz, but this assumption contradicts the definition of xPy as holding if and only if the subject chooses x over y whenever he must choose between only x and y It may well happen that he chooses x over y and y over z, but is never presented with a choice between x and z

Secondly, it may happen that on several occasions when the subject is presented with a choice between x and y, he sometimes chooses x and sometimes y In such a situation, neither xPy nor yPx would hold, according to the definition, and hence, if the fundamental assumptions are correct, it would follow that x is indifferent to y But if x were chosen over y 99 per cent of the time, it would seem unreasonable to assert that the subject was indifferent between x and y

One final objection to the suggested interpretation of the preference relation P is that there is no simple interpretation of the relations I and R corresponding to this interpretation of P It is assumed that a choice in which the only alternatives are x and y does not permit the possibility that neither of the alternatives are chosen, but if one is chosen, then, according to the definition of P, that alternative is preferred to the one which is not chosen Hence, the relation I cannot be defined in terms of observed choices

Another way of interpreting the relations P, I, and R is suggested by the fact that the choice between two alternatives x and y may not always be the same on different occasions when those alternatives are presented xPy may be defined to mean that x is chosen over y more than 50 per cent of the time in decisions involving those two alternatives only The relation I is analogously defined such that xIy means that x and y are each chosen 50 per cent of the time in a decision between those two This interpretation gets around the difficulty raised against the former interpretation of P in which xPy was defined to mean that x is *always* preferred to y It is, however, subject to the same logical difficulty alluded to previously in the case in which no choices between certain pairs of alternatives are observed Another important objection to this interpretation is that the strong *a priori* reasons which can be advanced as to why a subject should behave in accordance with Assumptions 1 and 2 (to be discussed in Section 5 on axioms for Bernoullian Utility theory) do not apply, or apply with less force, with this second interpretation of the preference relations

A third way of interpreting the relations P, I, and R is as representing subjective preferences at a given time This interpretation of course makes the theory useless as a description of overt behavior in decision situations, unless further hypotheses are added specifying the relation between subjec-

tive preference and observable behavior Many applications do, however, appear to make this interpretation of preference, and such a theory depends on certain additional assumptions as to how subjective preference can be inferred from observed actions These observations may be of several kinds, such as choices in decision situations, verbal reports, inferences from other choice data, etc We shall see below that this interpretation is the natural one to take if utility theory is interpreted *normatively*, that is, not as a description of actual behavior, but as a description of behavior which in a sense is *rational*, and which one ought to conform to if one hopes to achieve one's aims

3.5.3 Interpretations of the Utility Function.

Since utility values are thought of as representing strengths of preferences, and preference strengths are inaccessible to direct observation, the problem of giving utility an empirical significance is particularly important, if this theory is to be applied In the interpretation of utility, there are two major approaches The usual approach is to take Assumptions 1 and 2 as *definitions* of utility or, rather, to regard them as asserting that there *exists* a function u which has the relations to the set K and the relation R postulated in the assumptions The assumption then is that there exists a function u with domain K such that for all x and y in K, xRy if and only if

$$u(x) \geq u(y),$$

and for all $0 \leq p \leq 1$,

$$u(<px, (1-p)y>) = pu(x) + (1-p)u(y)$$

Once the interpretations of K and R are defined, the question of whether there is some function u which satisfies the above conditions is perfectly definite, and it becomes possible in some cases to state certain observable criteria which are necessary and sufficient to assure that the assumption is true It is this interpretation of utility which is tacitly assumed in most applications of utility theory We shall call any function u satisfying these conditions a "*Bernoullian Utility function*"

Definition of "Bernoullian Utility Function." A real valued function u with domain K which satisfies conditions 1 and 2 below is a Bernoullian Utility function

(1) For all x and y in K, xRy if and only if

$$u(x) \geq u(y)$$

(2) For all x and y in K, and all $0 \leq p \leq 1$,

$$u(<px, (1-p)y>) = pu(x) + (1-p)u(y).$$

Notice that this definition does not make reference to anything subjective, except insofar as the relation R may be defined subjectively. Nothing is said here about $u(x)$ being a numerical measure of the strength of preference; all that is required is that the function u satisfy the two conditions above. Utility is then interpreted as any function satisfying the conditions of the definition; *i. e.*, a utility function is defined to be a Bernoullian Utility function. Corresponding to this interpretation, of course, the fundamental assumptions must be modified. As Assumptions 1 and 2 are formulated it is tautological to assert that a utility function satisfies them, since the utility functions are defined in terms of them. They can be made empirically meaningful by reformulating them to assert that there is a Bernoullian Utility function:

Assumption A. There exists a Bernoullian Utility function.

Assumption A may be thought of as a weak form of the two original assumptions. What it amounts to is using these two assumptions to define the utility function which in turn determines the meaning of the assumptions.

If the values of the utility function could be determined by an independent method of measurement (say, one of the alternative methods discussed briefly in Section 2), the hypotheses of Bernoullian Utility theory would be stronger in predictive content than they are in the weak form stated in Assumption A. The reason for this fact is that under Assumption A, it is first necessary to observe some choices in order to gain some information about the utility values before these values can be used in predicting other choices, whereas if the values are already available through some other measurement, then predictions can be made immediately about preferences. Since, however, the interest in utility measurement is comparatively recent, there are in most cases no independent measures of the utility values of sets of alternatives available, and Bernoullian Utility theory must be given the weak interpretation if it is to be applied at all.

3.5.4 Descriptive vs. Normative Interpretations of Bernoullian Utility Theory.

We have discussed the concepts of Bernoullian Utility primarily as a theory of actual individual behavior in decision situations. As originally

formulated by von Neumann and Morgenstern, however, the theory was meant to be taken as *normative*, describing the behavior of a rational man in decision situations The contrast between a scientific or descriptive theory and a normative theory can be illustrated by the example of the discipline of logic The aim in logic is to describe correct, or rational, reasoning processes, not actual reasoning processes, which belong to the province of psychology From the point of view of psychology, the distinction between correct and incorrect reasoning is not of primary importance (except as incorrect reasoning may have consequences adversely affecting the organism), whereas this distinction is all important in logic Analogously, in a normative theory of decision making the aim is to determine what are intelligent ways of acting whether or not people do in fact act intelligently

The important distinction between the normative and descriptive interpretations of Bernoullian Utility theory is in the interpretations of the basic assumptions, since the interpretations of the basic concepts are similar in both the normative and descriptive theories In the case of the descriptive theory, the assumptions are tested by the conformity of predictions based on them with observed choices In the normative theory the assumptions are tested by their conformity to some criteria as to what constitutes rational behavior It will be seen in later sections that the basic assumptions of utility theory can be strongly justified in some instances as rules to which a rational person should conform, though there are reasons often why a person would not actually behave in accordance with the assumptions simply because he would be unable to think through the consequences of the acts among which he must choose

3.6 Uniqueness of Bernoullian Utility Functions

We shall show in this section that if a set of preferences satisfies the assumptions of Bernoullian Utility theory — if there exists a Bernoullian Utility for these preferences — then this function is uniquely determined once a zero point and unit of utility are arbitrarily chosen The simplest way to show this is simply to assume that two alternatives x_0 and x_1 are picked and arbitrarily assigned utilities 0 and 1, respectively (it must, of course, be assumed that x_1 is preferred to x_0) This stipulation fixes the zero point of utility as being the utility of alternative x_0 and the unit of utility as the difference in utility between x_1 and x_0 Now, to determine the utility of an arbitrary alternative x, three cases must be considered (1) x is preferred to x_1, (2) x_1 is preferred-or indifferent to x and x is preferred or indifferent to x_0, (3) x_0 is preferred to x In case (1), we have xPx_1 and x_1Px_0, and there

exists some probability p such that x_1 is indifferent to the mixture $<px, (1-p)x_0>$ In this case,

$$\begin{aligned} u(x_1) &= u(<px, (1-p)x_0>) \\ &= pu(x) + (1-p)u(x_0) \end{aligned}$$

Since $u(x_1) = 1$ and $u(x_0) = 0$, the above equation can be solved for $u(x)$ to yield

$$u(x) = \frac{1}{p}$$

In case (2), xRx_1 and xRx_0, and therefore there is some probability p such that x is indifferent to the mixture $<px_1, (1-p)x_0>$ Then

$$\begin{aligned} u(x) &= u(<px_1, (1-p)x_0>) \\ &= pu(x_1) + (1-p)u(x_0) \\ &= p \end{aligned}$$

In the last case, x_0Px, and there is some probability p such that x_0 is indifferent to the mixture $<px_1 (1-p)x>$ In this case, then,

$$\begin{aligned} u(x_0) &= u(<px_1, (1-p)x>) \\ 0 &= pu(x_1) + (1-p)u(x) \\ 0 &= p + (1-p)u(x) \end{aligned}$$

Therefore,

$$u(x) = \frac{-p}{1-p}$$

The foregoing proof shows in general that Bernoullian Utility is, in the classical terminology, a "cardinal utility measure' In more technical terms, utility (in this theory) is shown to be a *linear scale* (see Guilford, [1954], or Coombs, Raiffa, and Thrall [1954])

It is worthwhile to observe one rather unrealistic assumption inherent in this proof It is assumed that a unique probability p can be found which will make a certain mixture indifferent to another alternative The existence and uniqueness of this probability follows, of course, from the Bernoullian Utility assumptions, but is unrealistic from an experimental point of view Thus, there is likely to be a range of probabilities p such that a subject would

judge an alternative $\langle px_1, (1-p)x_0\rangle$ indifferent to x. The width of this range can be regarded as a measure of the subject's ability to discriminate between probabilities.

4 OBSERVABLE CONSEQUENCES OF THE ASSUMPTIONS OF BERNOULLIAN UTILITY THEORY

4.1 Consequences of the Ordinal Assumption

The object of this section is to derive some of the consequences which follow from the assumption that an individual has a Bernoullian Utility function (Assumption A). The consequences determine the empirical significance of the fundamental hypotheses formulated in the previous section. It would be impossible to derive all of the possible observable consequences of these assumptions, hence it is necessary to introduce some method of selection. The consequences to be discussed will be selected with two aims in view: first, they should reveal directly observable implications of the basic hypotheses, and second, they should be as *strong* as possible in the sense that all, or nearly all, of the implications of the fundamental hypotheses could also be derived from these consequences of the basic hypotheses. With the usual interpretation of the fundamental concepts of utility theory, only acts of choosing are directly observable, and preferences can be directly inferred from these choices whereas utilities cannot. Therefore, in order to reveal the empirical significance of the basic assumptions, it is necessary to determine what they imply for the individual's preferences. In this part, we shall derive the most important empirical consequences of the Ordinal Assumption (or rather, the weak form of the assumption that the individual *has* an ordinal utility function).

The Ordinal Assumption can be stated: there exists a real valued function u over the domain K such that for all x and y in K, xRy if and only if

$$u(x) \geq u(y)$$

Consequence 1, below, is the most important of the empirical consequences of the Ordinal Assumption, in that most of the observable consequences of the Ordinal Assumption also follow from Consequence 1.

Consequence 1. R is a *weak ordering* of the set K, i.e., R satisfies the following two conditions,

(1) for all x and y in K, either xRy or yRx,
(2) for all x, y, and z in K, if xRy and yRz, then xRz

It is not difficult to see that if an individual satisfies the Ordinal Assumption, then his preference-or-indifference relation R must conform to the two conditions of Consequence 1 First, suppose that x and y are two alternatives in the set K Then, if u is a utility function, it must be that either $u(x) \geq u(y)$ or $u(y) \geq u(x)$ (or possibly both, this possibility is admitted by the "inclusive" sense of the word 'or") According to the Ordinal Assumption, then, either xRy or yRx, or both, and therefore the first of the two conditions of Consequence 1 is proven

To prove the second of the conditions above, it must be shown that if xRy and yRz, and the individual satisfies the Ordinal Assumption, then xRz According to the Ordinal Assumption xRy implies $u(x) \geq u(y)$ and yRz implies $u(y) \geq u(z)$, and hence $u(x) \geq u(z)$ Conversely, however, $u(x) \geq u(z)$ implies xRz, and the second condition is proven

Now we may examine more closely the significance of the Ordinal Assumption by seeing what Consequence 1 says about preferences It is easier to do this by translating this Consequence into conditions which apply to the preference and indifference relations P and I directly, rather than to the preference-or-indifference relation These conditions follow from Consequence 1 and the definitions given in Section 3 of P and I in terms of R

Corollaries to Consequence 1.

1 1 For all x and y in K, exactly one of xPy, xIy, or yPx holds
1 2 For all x, y, and z in K, if xPy and yPz, then xPz
1 3 For all x in K, xIx
1 4 For all x and y in K, if xIy, then yIx
1 5 For all x, y, and z in K, if xIy and yIz, then xIz
1 6 For all x, y, and z in K, if xPy and yIz, then xPz, and if xPy and xIz, then zPy

The intuitive significance of these conditions can be divided into two parts First, the relation of preference must be *transitive* and *asymmetric*, if the individual prefers x to y and y to z, then he must prefer x to z, and if he prefers x to y, then he must not prefer y to x This is what one would expect simply from *a priori* considerations, particularly if his choices are made during a short interval of time, so that his preferences remain substantially fixed in the interval These assertions follow from conditions 1 1 and 1 2 Conditions 1 3 — 1 6 imply that, for the purposes of comparisons of prefer-

ence, two alternatives which are indifferent are regarded as identical. If x is indifferent to y, and y has any relation of preference to z, then conditions 1.3 — 1.6 imply that x has the same relation to z.

Certain of conditions 1.3 — 1.5, applying to the indifference relation, would not be expected on *a priori* grounds. If xIy is interpreted to mean that x is not noticeably different from y in subjective value, one would not expect conditions 1.5 and 1.6 to hold. It might well be that x is not noticeably different from y, and y is not noticeably different from z, but nevertheless x is preferred to z, which would contradict condition 1.5 requiring that x be indifferent to z. A similar argument shows that a case might arise in which x would be preferred to y, and y would be indifferent to z, but x would not be preferred to z, contradicting condition 1.6.

The counter-intuitive nature of conditions 1.5 and 1.6, and the fact that it is usually found that when a subject does not satisfy the Ordinal Assumption (it is conditions 1.5 and 1.6 which are violated), have led many authors to suggest the need for a revision of the theory (or that part of it contained in the Ordinal Assumption) to bring it into conformity with observed behavior. Against this demand for revision it is argued that Consequence 1 gives a good, though obviously not perfect, approximation to observed behavior, and that the gain in accuracy resulting from revision would have to be paid for at a cost of greatly increased mathematical complexity. We shall, however, examine some proposals for dealing with intransitive indifferences in Section 5.

In many cases Consequence 1 includes all the possible observable consequences of the Ordinal Assumption, since all other consequences can also be derived from Consequence 1. It can be shown, for example, that if the set A contains only a finite number of alternatives, or even a denumerable infinity (a set is denumerably infinite if its members can be paired off with the integers) of alternatives, then Consequence 1 not only is implied by the Ordinal Assumption, but implies it, and hence all consequences which follow from the Ordinal Assumption follow from Consequence 1. The Ordinal Assumption and Consequence 1 are no longer equivalent if A contains a larger number of alternatives, and therefore there will be consequences derivable from the Ordinal Assumption which are not derivable from Consequence 1 for such sets. Owing to the fact that in the general theory of Bernoullian Utility it is possible to form from two alternatives x and y a different mixture alternative $<px, (1-p)y>$ for each real number p between 0 and 1, there is more than a denumerable infinity of alternatives in any mixture space, and therefore the Ordinal Assumption and Consequence

1 are not equivalent for sets of alternatives which are mixture spaces. However, no one has as yet investigated the implications of the Ordinal Assumption beyond Consequence 1, and it may be wondered what empirical significance such implications might have which are only revealed in non-denumerably infinite sets of alternatives. Debreu [1954] has investigated the converse problem of determining sets of conditions on the preference relations which imply the existence of an ordinal utility function.

4.2 Consequences of the Mixture Space Axioms.

In the previous section we derived the consequences of the Ordinal hypotheses alone. These consequences are part of all utility theories which incorporate the Ordinal Assumption, hence belong to all the present utility theories with the exception of Luce's theory of discrimination structures. The mixture space axioms (in some instances weakened or modified) are likewise common to most theories of the utility of mixture alternatives. Therefore it is worthwhile to develop the consequences of them in a separate section. In the deduction of these consequences, we shall always assume the Ordinal hypothesis as well, and hence make use of Consequence 1 and its corollaries.

The essential significance of the mixture space axioms is that they assert identities between various mixture alternatives. For example, Axiom M2 asserts that the two alternatives $<px, (1-p)y>$ and $<(1-p)y, px>$ are equal. If the concept of "equality" is to have its usual logical sense, then it should be the case that if any alternative x is equal to another alternative y, and y has any preference relation to a third alternative z, then x should have the same relation to z. Since Consequence 1 implies that any two alternatives which are indifferent must have the same preference relation to any third alternative, the essential content of the mixture space axioms can be expressed in the assumption that any two alternatives which are equal (according to the mixture space axioms) are indifferent. This conclusion is formulated in Consequence 2.

Consequence 2. For all x and y in K (assumed to be a mixture space), if $x = y$, then xIy.

The important corollaries of Consequence 2 can be obtained immediately from the Axioms M2, M2, and M4 given in Section 3 for mixture spaces.

Corollaries to Consequence 2.

2.1 For all x and y in A and $0 \leq p \leq 1$,

$$\langle px, (1-p)y\rangle \, I \, \langle (1-p)y, px\rangle$$

2.2 For all x, y, and z in A and $0 \leq p \leq 1$ and $0 \leq q \leq 1$ such that p and q are not both zero,

$$\langle px, (1-p)\langle qy, (1-q)z\rangle\rangle \, I$$

$$\langle (p+q-pq)\left\langle \frac{p}{p+q-pq}x, \frac{q-pq}{p+q-pq}y\right\rangle, (1-p)(1-q)z\rangle.$$

2.3 For all x in A and all $0 \leq p \leq 1$,

$$\langle px, (1-p)x\rangle \, Ix$$

These corollaries to Consequence 2 show that the Axioms for mixture spaces are not trivial as far as their empirical implications are concerned. Corollaries 2.1 and 2.3 are fairly innocuous it is true. Corollary 2.1 asserts that an individual will be indifferent between the alternative of getting x with probability p or else getting y, and the alternative of getting y with probability $1-p$, otherwise getting x. Hardly any reflection would be required to make the individual realize that these two alternatives are essentially the same, since each involves getting the outcome x with probability p and y with probability $1-p$. Corollary 2.3 is likewise intuitively evident: it says only that the individual is indifferent between the alternative of getting x with probability p, or else getting x, and the alternative x. Corollary 2.2 cannot, however, be dismissed as trivial. To take a specific example, suppose that the probabilities p and q are both 0.5, then 2.2 asserts that the mixture

$$\langle .5x, .5\langle .5y, .5z\rangle\rangle$$

is regarded as indifferent to the mixture

$$\langle .75\langle .667x, .333y\rangle, .25z\rangle$$

It is true that computing the probabilities of getting the outcomes x, y, and z in both mixtures gives a probability 0.5 of getting x, 0.25 of getting y, and 0.25 of getting z, and therefore both alternatives yield the same outcomes with the same probabilities. However, one may very well question whether these two alternatives appear the same to the subject. At the very least, it seems to be asking a great deal of the subject's knowledge of probability to expect that he should be able to recognize that the two alterna-

tives above are essentially the same A second difficulty arises when the conclusions of 2 2 and 2 3 are combined According to 2 2, the alternative

$$< 5x, < 5x, 5y>>$$

must be indifferent to

$$< 75 < 667x, 333x>, 25y>$$

According to Axiom M4, $< 667x, 333x>$ is equal to the alternative x, and therefore it follows that

$$< 5x, < 5x, 5y>> I < 75x, 25y>$$

In this case the probabilities of the outcomes involved in both alternatives are the same, but one alternative involves the taking of two risks, whereas the other involves taking only one Therefore it might be argued that subjects who reacted either positively or negatively to the taking of risks *per se* would not regard the two alternatives as equally preferable

Against these objections to Corollary 2 2 it may be replied that if one is to develop a theory of choices among risk alternatives in which the utilities of the alternatives depend only on the possible outcomes and their relative probabilities, then that theory must automatically presuppose that any two alternatives with the same outcomes and the same probabilities must have the same utility, and hence be regarded as indifferent Therefore, if one desires to develop a theory which will meet the objections raised against Corollary 2 2, that theory will necessarily have to take into account more factors in the subject's judgment of alternatives than simply their possible outcomes and their probabilities It is to be observed, however, that while most of the experimental work which has so far been done has either presupposed Bernoullian Utility theory (and hence the mixture space axioms and their consequences) or else has been directed at testing the hypotheses of Bernoullian Utility theory, most of these experimental tests have avoided the objections raised above by confining their observations to comparisons between alternatives in which the probabilities involved are simple and only one risk is involved in a given alternative

4.3 Consequences of the General Hypotheses of Bernoullian Utility Theory.

The empirical content of the hypotheses of Bernoullian Utility theory, beyond what is contained in Consequences 1 and 2, can be formulated in three conditions on the relations P and I It can be shown that these condi

tions, together with Consequences 1 and 2, are actually equivalent to Assumption A, although the proof of this fact is too difficult to be included here

Consequence 3.

3 1 For all x, y, and z in Λ and $0 \leq p \leq 1$, if xIy, then
$<px, (1-p)z> I <py, (1-p)z>$

3 2 For all x, y, and z in Λ and $0 < p \leq 1$, if xPy, then
$<px, (1-p)z> P <py, (1-p)z>$

3 3 For all x, y, and z in Λ, if xPy and yPz, then for some
$0 < p < 1$ and some $0 < q < 1$, $<px, (1-p)z> Py$, and
$yP < qx, (1-q)z>$

It is important to note the cases in the above consequences in which the strict inequality signs replace the usual '$\leq$' signs

Consequences 3 1 — 3 3 follow immediately from the definition of a Bernoullian Utility function, and the assumption that there exists a Bernoullian Utility function for the set Λ (Assumption A) To prove 3 1, suppose that xIy and u is a Bernoullian Utility function Then it follows that $u(x) = u(y)$, and

$$u(< px, (1-p)z>) = pu(x) + (1-p)u(z)$$

$$= pu(y) + (1-p)u(z)$$

$$= u(< py, (1-p)z>),$$

and therefore $< px, (1-p)z > I < py, (1-p)z >$ To prove 3 2, suppose that xPy and $0 < p \leq 1$ Then

$$u(x) > u(y)$$

$$pu(x) > pu(y)$$

$$pu(x) + (1-p)u(z) > pu(y) + (1-p)u(z)$$

$$u(<px, (1-p)z >) > u(< py, (1-p)z>),$$

and therefore $< px, (1-p)z > P < py, (1-p)z >$ Finally, to prove 3 3, suppose that xPy and yPz, then $u(x) > u(y) > u(z)$ Now it must be that

$$u(x) - u(z) > u(y) - u(z) > 0,$$

hence

$$1 > \frac{u(y) - u(z)}{u(x) - u(z)} > 0.$$

Then it is possible to pick numbers p and q such that:

$$1 > p > \frac{u(y) - u(z)}{u(x) - u(z)} > q > 0.$$

Now, if the mixtures $< px, (1-p)z >$ and $<qx, (1-q)z >$ are formed, it follows immediately that:

$$u(< px, (1-p)z >) > u(y) > u(<qx, (1-q)z >).$$

Therefore $<px, (1-p)z > Py$ and $yP < qx, (1-q)z >$, and 3.3 is proven.

Consequence 3 is very important from the point of view of the observable behavior it implies. In order to get a better idea of its implications, we shall state some of its corollaries.

Corollaries to Consequence 3.

3.4 If xPy, then for all $0 < p < 1$, $xP < px, (1-p)y>$, and

$$<px, (1-p)y > Py.$$

3.5 If xPy and $1 \geq p > q \geq 0$, then

$$< px, (1-p)y> P < qx, (1-q)y>.$$

3.6 If xPy and yPz, then there exists $1 > p > 0$ such that

$$<px, (1-p)z)> Iy.$$

Corollary 3.4 follows from 3.2 (setting $z = y$) and from 2.3 (since, according to 2.3, $<py, (1-p)y > Iy$). Corollary 3.5 follows from 3.4 and Consequence 2, since

$$< qx, (1-q)y > = <\frac{q}{p}<px, (1-p)y>, (1-\frac{q}{p})y >.$$

According to 3.4, $<px, (1-p)y> Py$, and therefore

$$< px, (1-p)y > P <\frac{q}{p}px, (1-p)y, (1-\frac{q}{p})y >.$$

Corollary 3 6 follows from 3 3 The proof depends on the ordinal properties of real numbers and will be omitted here

It might seem on superficial examination that Corollary 3 4 is intuitively evident since it represents what any rational person would do This corollary says that if a person prefers x to y, then he prefers x to any mixture $<px, (1-p)y>$ in which there is some chance of getting y, and likewise he prefers this mixture in which there is some chance of getting x to the certainty of getting y This can be interpreted as a special case of what L J Savage [1954] has called the 'sure thing principle In this case, the mixture $<px, (1-p)y>$ ought to be preferred to y, since the only possible outcomes of the mixture are x and y, either of which is at least as good as y, and one of which (the alternative x) is definitely preferred to y Hence, no matter what the outcome of the risk alternative $<px, (1-p)y>$ is, it will be at least as good as the result for accepting y for certain, and there is a chance it will be better Similarly, x ought to be preferred to the mixture $<px, (1-p)y>$ since no outcome of the mixture is better than x, and there is a chance that the outcome of the mixture will be y, which is definitely worse than x

In spite of the apparently conclusive arguments that a rational person should act in accordance with Corollary 3 4, there are seemingly instances of behavior which violates this rule Marschak [1950] has cited the example of mountain climbers who appear to prefer climbing and therefore taking a certain risk of serious injury or death to not climbing and thus avoiding this risk Here, the alternative 'living safely corresponds to x in 3 4, and 'being seriously injured correspond to y It would seem certain that x is preferred to y, and hence, if the subject conforms to Corollary 3 4, x should be preferred to any mixture $<px, (1-p)y>$ (where p is less than 1) However, the mountain climber prefers $<px, (1-p)y>$ (corresponding to climbing and taking a risk $1-p$ of being killed) to x, therefore violating Corollary 3 4

An even more extreme example of behavior violating 3 4 arises in the "game" of Russian Roulette, in which the "player places a revolver with one chamber loaded against his head, spins the chamber until it stops, and then pulls the trigger, killing himself if the chamber has stopped with the bullet in firing position If the revolver is a six shooter, the Russian Roulette player takes a chance of 1/6 of killing himself each time he plays If x represents the outcome of staying alive, and y represents the outcome of being killed, then the alternative of playing the game once should be represented by the mixture $<5/6x, 1/6y>$ If, as seems reasonable to assume,

the player would prefer the certainty of x (staying alive) to the certainty of y (being killed), then he should also prefer the certainty of x to the mixture $<5/6x,\ 5/6y>$ according to 3 4 But Russian Roulette players seem to prefer to play (at least once anyway) and take the risk rather than following the safe course

The examples given above, as well as many more mundane ones in which a person seems to prefer taking a risk rather than following a safe course, can be explained in such a way as to 'save the principle" of Corollary 3 4 In the case of the mountain climber, it is certainly an over-simplification to view his alternatives as having only two possible outcomes staying unhurt, or being seriously injured It is tacitly assumed that the x (staying uninjured) involved in the alternative of not going climbing is identical with the x involved in going climbing However, it is obviously not the same to a mountain climber whether he stays at home and avoids injury or goes climbing and avoids injury, yet these two "outcomes" were assumed to be the same in the above argument against 3 4 The more bizarre case of the Russian Roulette player can be explained in a similar way the outcomes of "staying alive which are involved in both the alternatives of not playing (alternative x) and playing (alternative $<5/6x,\ 1/6y>$) are not the same to him, since he presumably believes that there is something to be gained (in the prestige accruing from taking the risk) if he takes the risk and lives which he would not have if he simply refused to take the risk

Two observations are to be drawn from the above examples apparently contradicting Corollary 3 4 (and therefore the fundamental assumptions of Bernoullian Utility theory) and their explanations First, it becomes clear that one may be led to absurdities if the set K of alternatives and their outcomes is not carefully defined Thus, in the case of the mountain climber, the definition of K in terms solely of the two outcomes of 'being uninjured' and 'being seriously injured and probability mixtures of these led to the assumption that the two outcomes 'going climbing and being uninjured' and "staying at home and being uninjured were the same, and hence had the same subjective value To avoid these absurdities one must stipulate that the set K should be defined in such a way that no alternative should cover two different cases which have widely differing subjective values Secondly, it becomes clear that Bernoullian Utility theory is not applicable where the act of gambling in a probability mixture influences the value of the outcomes of the risk In the case of Russian Roulette, the act of taking the risk increases the value of the outcome of staying alive, since the players believe that gambling and winning is more valuable than not gambling

A strict adherence to the proviso that Bernoullian Utility theory cannot be applied in situations in which the act of taking a risk influences the value of the outcomes would logically rule out practically all applications of the theory, since almost any risk probably influences the subjective values of its outcomes in some degree Like all psychological theories, Bernoullian Utility theory is an approximation, and the important question is whether it is sufficiently accurate to be useful in a particular situation What the above arguments show is that the theory is likely to be more inaccurate in situations in which risk taking has high intrinsic value (or has a large effect on the values of its outcomes) than in situations in which its intrinsic value is low Therefore, fruitful applications are to be looked for in situations in which the act of taking a risk has little influence on the value of its outcomes

Corollary 3 6 has, like 3 4, some counter instances Behavior apparently violating 3 6 occurs where the three alternatives x, y, and z are highly disparate in value Thrall [1954] gives the following example $x =$ be given two pins, $y =$ be given one pin, and $z =$ be hanged at sundown It seems reasonable to suppose that x is preferred to y and y is preferred to z, xPy and yPz According to 3 6 then, there should exist a probability $0 < p < 1$ such that the alternative $< px, (1-p)z >$ is indifferent to y, i e, the alternative of receiving two pins with probability p or being hanged with non zero probability $1-p$ is regarded as indifferent to the alternative of receiving one pin But many people would probably feel that if there were any chance at all of their being hanged then they would prefer to take the certainty of getting one pin to the risk which includes this possibility of hanging For these people y would be preferred to the mixture $< px, (1-p)z >$, no matter how small the probability $1-p$ is, as long as it is greater than zero This behavior would contradict Corollary 3 6

Closer examination of the above example reveals the difficulty inherent in attempting to give psychological meaning to the probabilities involved in the probability mixtures If there were any probability in the example such that the mixture $< px, (1-p)z >$ were indifferent to the sure alternative y, it would certainly be very close to 1, so that the probability $1-p$ of being hanged would be very small It is doubtful though whether any psychological significance can be attached to a probability of say one millionth of being hung If a person were to say that he would still prefer to take the single pin rather than the risk involving a probability of one millionth of being hung, one could then go on to ask him how he felt about the risk alternative if the probability of being hung were reduced to a trillionth The

person might reject this risk also, though if it were pointed out that there would be considerably more danger of his being struck by lightning at sundown than being hung at sundown if he took the risk, he might change his mind What the example shows is that the psychological notion of probability is very hazy, and it is probably unrealistic to expect the theory to apply either where the probabilities involved are very close to zero, or where the subjective values of the alternatives are highly disparate

It is possible to follow two courses with respect to the difficulties raised in the last example Either it may be assumed that the theory is as it stands a good enough approximation to actual behavior to be useful, or it may be modified to allow for the possibility of alternatives of disparate or non-comparable subjective values We shall be primarily concerned with the theory of Bernoullian Utility as formulated in Assumption A, and therefore shall assume Corollary 3 6, even though actual behavior may contradict it in certain instances The other course has been followed by M Hausner and J G Wendel (Hausner, M and Wendel, J G [1952], Hausner, M [1954], see also Thrall, R. [1954]) who have developed a modified theory of utility in which Consequence 3 3 (and hence Corollary 3 6) does not hold We shall give a brief description of this theory in Section 5

4.4 Consequences of the General Bernoullian Utility Theory Assumptions for Mixture Spaces of Arbitrary Finite Numbers of Alternatives.

Assumption A is generalized to apply to mixtures of arbitrary numbers of alternatives as follows

Assumption A′. There exists a real valued function u over the set K such that

(1) for all x and y in K, xRy if and only if $u(x) \geq u(y)$,
(2) for all $x_1, \ldots, x_n$ in K and all non negative $p_1, \ldots, p_n$ such that

$$p_1 + p_2 + \cdots + p_n = 1,$$
$$u(<p_1x_1, \ldots, p_nx_n>) = p_1u(x_1) + \cdots + p_nu(x_n)$$

Consequences 1 and 2 follow from Assumption A′ in the same way as they follow from Assumption A, though the meaning of Consequence 2 is somewhat different. The definition of equality, on which Consequence 2 depends, as generalized to apply to general finite mixtures has the same

intuitive significance as applied to mixtures of two alternatives, but the axioms necessary to characterize it are more complicated. Intuitively, two mixture alternatives are defined to be equal if and only if each yields the same outcomes with the same probabilities.

Definition (Finite Mixture Spaces) A set K which statisfies Axioms M'1 — M'4 is a finite mixture space.

M'1 For all $x_1, \ldots, x_n$ (n arbitrary) in K, and all non negative $p_1, \ldots, p_n$ such that $p_1 + \cdots + p_n = 1$, $<p_1x_1, \ldots, p_nx_n>$ is an element of K.

M'2 For all $x_1, \ldots, x_n$ in K,
$x_1 = <1x_1, 0x_2, 0x_3, \ldots, 0x_n>$

M'3 For all $x_1, \ldots, x_n, y_1, \ldots, y_m$ in K and non negative numbers p_{ij}, and q_i, $i = 1, \ldots, n$ and $j = 1, \ldots, m$ such that $p_{i1} + \cdots + p_{im} = 1$,
and
$q_1 + \cdots + q_n = 1$,
if
$x_i = <p_{i1}y_1, \ldots, p_{im}y_m>$, $i = 1, \ldots, n$,
then
$<q_1x_1, \ldots, q_nx_n> = <r_1y_1, \ldots, r_my_m>$,
where
$r_j = q_1p_{1j} + \cdots + q_np_{nj}$, $j = 1, \ldots, m$

M'4 For all x, y, and z in K and for $0 < p \leq 1$, if
$< px, (1-p)z> = <py, (1-p)z>$, then $x = y$

It is easy to show that Axioms M'1 — M'4 imply the original mixture space axioms in the case in which probability mixtures of two alternatives only can be formed. Now, it can be shown that Consequence 2 follows from Assumption A' with the notion of equality of mixtures generalized to apply to arbitrary finite mixtures.

Consequence 3 also applies to the general finite mixture spaces, and with the same meaning as previously, and it can be shown that these three consequences are equivalent to the Assumption A' in that Assumption A' implies Consequences 1 — 3, and these consequences imply the assumption. The proof of this equivalence is difficult in the general case, and so we simply state it. In the following section, we shall show how it is possible to derive Assumption A' from its consequences in a special case in which each alternative is equal to a probability mixture of a fixed basic set of 'sure' alternatives (this set being assumed finite).

4.5 Axiomatization of Bernoullian Utility Theory: Representation Theorems.

In many discussions Bernoullian Utility theory has been presented axiomatically. This means that a set of axioms are given specifying certain conditions which must be satisfied by the preference relation, and such that these conditions imply Assumption A, *i.e.*, the axioms imply the existence of a Bernoullian Utility function. Von Neumann and Morgenstern [1947] gave a set of axioms in their original formulation of Bernoullian Utility, and many other writers on utility have explored the significance and implications of other systems of axioms (see, for instance, Marschak [1950], Friedman and Savage [1952], Samuelson [1952], Savage [1952]). The importance of developing the theory axiomatically lies in the fact that, if the axioms all have a more or less immediate empirical significance, the empirical meaning of the theory is much more clear than it would be if Assumption A were simply stated without justification.

We shall not examine the various axiomatic systems for the classical Bernoullian Utility theory here. It has been stated that Consequences 1 — 3 are actually equivalent to Assumption A, and therefore they could be taken as axioms for the theory of Bernoullian Utility. In order to prove that these axioms are actually *sufficient* (that is, that they imply Assumption A), it is necessary to prove what is called a "*representation theorem*." A representation theorem in this instance would have the form: if the set K and the relation R satisfy the given axioms, then there exists a Bernoullian Utility function for this system.

To illustrate the idea of a representation theorem, we shall add one axiom to the conditions contained in Consequences 1 — 3 (as applied to finite mixture spaces), and show that if these axioms are satisfied, then there exists a Bernoullian Utility function which so to speak "represents" the preference relation. The axiom to be added specifies that all alternatives are equal to mixtures of a fixed, finite set of basic "sure" alternatives.

Representation Theorem. Let K be a finite mixture space, and R be a relation over K satisfying the conditions of Consequences 1 — 3, and following condition:

there exist elements $a_1, \ldots, a_n$ in K such that for all x in K there exist non negative numbers $p_1, \ldots, p_n$ such that

$$x = < p_1a_1, \ldots, p_na_n >$$

Then there exists a Bernoullian Utility function for the set K.

To prove the representation theorem it is necessary actually to construct a real valued function u defined over A and then show that this function is a Bernoullian Utility function. It may be assumed without loss of generality that the basic set of alternatives are arranged in order of preference

$$a_1 R a_2,\ a_2 R a_3,\ \ldots,\ a_{n-1} R a_n$$

(the fact that they can be so arranged follows from the fact that the preference or indifference relation is a weak ordering). Now, two cases are possible: either $a_1 P a_n$ or $a_1 I a_n$. If $a_1 I a_n$, then it follows from the fact that R is a weak ordering that all of the alternatives $a_1, \ldots, a_n$ are indifferent, and hence by Consequences 2 and 3, all the alternatives in A are indifferent. This case is trivial, since the function u can be defined such that $u(x) = 0$ for all x in A, and this is easily seen to be a Bernoullian Utility function for this set of alternatives.

In the second case, $a_1 P a_n$, and for each a_i, $a_1 R a_i$ and $a_i R a_n$. Now, we wish to show that for each i there exists a number $0 \leq q_i \leq 1$ such that

$$a_i I < q_i a_1, (1-q_i) a_n >; \tag{1}$$

i.e., each a_i is indifferent to a mixture of the two alternatives a_1 and a_n. For any given i there are three possibilities: either $a_1 I a_i$ and $a_i P a_n$, $a_1 P a_i$ and $a_i P a_n$, or $a_1 P a_n$ and $a_i I a_n$. If $a_1 I a_i$, define $q_i = 1$, then, by Axiom M 2,

$$< q_i a_1, (1-q_i) a_n > = < 1 a_1, 0 a_n > = a_1,$$

and since a_i was assumed indifferent to a_1, it follows from Consequence 2 (that equal alternatives are indifferent) that

$$a_i I < q_i a_1, (1-q_i) a_n >$$

In the case that $a_1 P a_i$ and $a_i P a_n$, it follows from Corollary 3.6 to Consequence 3 that there exists a number $0 < q_i < 1$ such that

$$a_i I < q_i a_1, (1-q_i) a_n >$$

Finally, if $a_i I a_n$, define $q_i = 0$. In this case,

$$< q_i a_1, (1-q_i) a_n > = < 0 a_1, 1 a_n > = a_n,$$

and therefore

$$a_i I < q_i a_1, (1-q_i) a_n >$$

With the numbers $q_1, \ldots, q_n$ defined so as to satisfy the indifference relation (1), the function u can be defined. It was assumed that for any alternative x, there are some numbers $p_1, \ldots, p_n$ such that

$$x = \langle p_1 a_1, \quad , p_n a_n \rangle \tag{2}$$

$u(x)$ is now defined by setting

$$u(x) = q_1 p_1 + q_2 p_2 + \quad + q_n p_n \tag{3}$$

It will be seen further on that if there were two sets of numbers $p_1, \quad , p_n$ which satisfied the equation in (2), the sum on the right hand side of equation (3) would be the same for both, and therefore it would not make any difference which one of these sets of numbers were picked in defining the value $u(x)$

The function u having been defined, it remains to be shown that it is a Bernoullian Utility function To show this, it is first necessary to prove that xRy if and only if $u(x) \geq u(y)$ for any x and y in K Suppose that xRy, and suppose that the numbers $p_1, \quad , p_n$ are the ones by which $u(x)$ is defined in equations (2) and (3) and that the numbers $r_1, \quad , r_n$ satisfy the analogous equations for the alternative y i e ,

$$x = \langle p_1 a_1, \quad , p_n a_n \rangle, \tag{4}$$

$$y = \langle r_1 a_1, \quad , r_n a_n \rangle, \tag{5}$$

$$u(x) = q_1 p_1 + q_2 p_2 + \quad + q_n p_n, \tag{6}$$

$$u(y) = q_1 r_1 + q_2 r_2 + \quad + q_n r_n \tag{7}$$

Now, it will be recalled that the numbers $q_1, \quad , q_n$ were defined so that

$$a_i I \langle q_i a_1, (1-q_i) a_n \rangle$$

It follows from Consequences 3 1 (generalized to finite mixture spaces) that if the alternatives $\langle q_i a_1, (1-q_i) a_n \rangle$ *are substituted for the alternatives* a_i, which they are indifferent to, in $\langle p_1 a_1, \quad , p_n a_n \rangle$ and $\langle r_1 a_1, \quad , r_n a_n \rangle$, the resulting alternatives will be indifferent to x and y, respectively i e , if

$$x' = \langle p_1 \langle q_1 a_1, (1-q_1) a_n \rangle, \quad , p_n \langle q_n a_1, (1-q_n) a_n \rangle \rangle \tag{8}$$

and

$$y' = \langle r_1 \langle q_1 a_1, (1-q_1) a_n \rangle, \quad , r_n \langle q_n a_1, (1-q_n) a_n \rangle \rangle, \tag{9}$$

then xIx' and yIy' According to Axiom M 3 in the definition of a finite mixture space,

$$x' = \langle (p_1 q_1 + \quad + p_n q_n) a_1, (p_1(1-q_1) + \quad + p_n(1-q_n)) a_n \rangle, \tag{10}$$

and

$$y' = \langle (r_1 q_1 + \quad + r_n q_n) a_1, (r_1(1-q_1) + \quad + r_n(1-q_n)) a_n \rangle \tag{11}$$

According to the definition of $u(x)$ and $u(y)$, it is easily seen that the above two equations are the same as

$$x' = \langle u(x)a_1, (1-u(x))a_n \rangle, \tag{12}$$

and

$$y' = \langle u(y)a_1, (1-u(y))a_n \rangle \tag{13}$$

Now, if xRy and xIx' and yIy, then it follows from Consequence 1 that $x'Ry$ To show that $u(x) \geq (u)y$, suppose that this is false, and $u(y) > u(x)$ Then, according to Corollary 3 5 and equations (12) and (13), it would follow that $y'Px$ (since a_1Pa_n), contradicting the supposition that xRy Conversely, if $u(x) \geq u(y)$, then by the same corollary it follows that $x'Ry$ and hence xRy

It has now been shown that the function u defined satisfies one of the two conditions necessary for a Bernoullian Utility function This proof incidentally shows that equation (3) is a definition, and that if there were two different sets of numbers $p_1, \quad, p_n$ satisfying equations (2) and (3), they would yield the same sum on the right of equation (3) If the sums yielded were different, it would be possible to use the same argument to show that $u(x) > u(x)$ and hence xPx, contrary to Consequence 1

It now remains to prove that $u(x)$ satisfies the second criterion for a Bernoullian Utility function if

$$p_1 + p_2 + \quad + p_n = 1$$

and the p_i's are non negative, then

$$u(\langle p_1x_1, \quad, p_mx_m \rangle) = p_1u(x_1) + p_2u(x_2) + \quad + p_m u_m(x_m)$$

For each of the x_i's there are numbers $r_1^i, \quad, r_n^i$ such that

$$x_i = \langle r_1^i a_1, \quad, r_n^i a_n \rangle \tag{14}$$

Then, according to the definition of u,

$$u(x_i) = q_1 r_1^i + q_2 r_2 + \quad + q_n r_n^i \tag{15}$$

According to axiom M 3, however,

$$\langle p_1 x_1, \quad, p_m x_m \rangle = \langle s_1 a_1, \quad, s_n a_n \rangle, \tag{16}$$

where

$$s_j = p_1 r_j^1 + p_2 r_j^2 + \quad + p_n r_j^n \tag{17}$$

$u(<p_1x_1, \quad , p_mx_m>)$ is therefore

$$u(<p_1x_1, \quad , p_mx_m>) = q_1s_1 + q_2s_2 + \quad + q_ns_n \tag{18}$$

It can easily be shown (by rearranging the terms in the above sums) that

$$\begin{aligned} u(<p_1x_1, \quad , p_mx_m>) &= p_1[q_1r_1^1 + \quad + q_nr_n^1] \\ + p_2[q_1r_1^2 + \quad + q_nr_n^2] + \quad &+ p_n[q_1r_n^1 + \quad + q_nr_n^n], \\ &= p_1u(x_1) + p_2u(x_2) + \quad + p_nu(x_n) \end{aligned} \tag{19}$$

Thus, the proof that $u(x)$ is a Bernoullian Utility function is concluded

5 MODIFICATIONS OF CLASSICAL BERNOULLIAN UTILITY THEORY

5 1 Preliminary Remarks.

Since the original appearance of the *Theory of Games and Economic Behavior*, a number of important modifications of the 'classical Bernoullian Utility theory have been proposed, mostly motivated by the necessity for bringing the structure of the classical model into closer conformity with the actual behavior which it is supposed to describe Let us state briefly again what some of the most obvious defects of the classical model as a descriptive theory are Perhaps the most fundamental defect of the theory is shared with all other utility theories namely, it requires perfect and consistent discrimination of subjective values and allows no "margin of error " As an idealization this assumption of perfect discrimination is innocuous enough, and is no worse than comparable assumptions to be found everywhere in the physical sciences, but it is fatal when it comes to making an experimental test of the theory, since subjects almost never fit the theory exactly One type of modification in the direction of empirical realism, therefore, is the replacement of the assumption of perfect discrimination (partly represented by the assumption that the indifference relation is transitive) by something weaker Two more defects of the classical theory are closely related They are (1) that a subject is aware of what the true or objective probabilities are in decision situations, and (2) that he is a 'perfect computer who is able to work out the probabilities of mixtures of mixtures in terms of the basic probabilities of the component mixtures The first of these difficulties may be met in two ways in an experimental situation either by training the

subjects so that in some sense they become "aware" of what the true probabilities are, or by assuming that the subjects do not necessarily react to the objective probabilities but instead to a subjective probability which might be thought of as what the subject "perceives" the objective probability to be. The second difficulty may be overcome experimentally by restricting the set of alternatives to mixtures of the first order, and not including mixtures of mixtures. One last difficulty inherent in the classical theory is that it takes no account of the "utility of gambling" (such as occurred in the Russian Roulette example).

In this section we shall discuss several modifications which aim to overcome the defects listed above. Many of these modifications are complicated, and an adequate discussion of them would have to be almost as extensive as our exposition of the classical theory. Limitations of space prevent this extensive treatment, and all that can be done here is to indicate briefly for each one what its fundamental assumptions are and how it is related to the classical theory.

5.2 Multi-dimensional Utilities.

The first modification to be discussed aims to cure a relatively minor "defect" in the classical assumptions. This is the assumption that probability mixtures satisfy the "Archimedian" condition (Corollary 3.6) that for any three alternatives x, y, and z such that xPy and yPz there must be some probability p such that $<px, (1-p)z>$ is indifferent to y. One counter-instance which apparently violates this assumption has been mentioned in Section 4. This is the situation in which the difference in utility between x and y is "incomparably smaller" than the difference between y and z, for example, where x is the alternative of getting two pins, y is getting one pin, and z is being hanged at sundown. Melvin Hausner (Hausner, M. [1954], see also Thrall, R. M. [1954], and Hausner, M. and Wendel, J. G. [1952]) has developed a theory of mixture spaces ordered by a preference relation which does not satisfy Consequence 3.6, which states formally the Archimedian condition.

We have stated that the hypotheses of Bernoullian Utility theory actually follow from the three empirical consequences stated in Section 4. If Bernoullian Utility theory is formulated axiomatically, then these consequences or those equivalent to them are taken as axioms, and the main deduction from the axioms is the "representation theorem" showing that if these axioms are satisfied, then a Bernoullian Utility function exists. Naturally, if the Archimedian assumption of Consequence 3.6 is dropped, then it is no

longer possible to prove that a Bernoullian Utility function exists However, it can be shown that if this assumption is omitted, then preferences can be represented by a more complicated sort of utility, which may be thought of as a generalization of the kind we have been dealing with up to this point The intuitive idea is that in the non-Archimedian case the strength of preference for an alternative is not represented by a single number, as it is in the ordinary theory, but by a *vector* A clearer conception of this representation can be gained from an example

Suppose that a member of a man's family is ill and needs some medicine The man has a choice of several kinds of medicine, which will be of varying degrees of effectiveness in curing the illness, and of varying costs Assuming that the costs of the medicines are not prohibitively high for the man, it is reasonable to suppose that he will choose among the medicines on the basis of their degrees of effectiveness in curing the illness, and only in case two were equally effective in this respect, would he take into account their relative costs in choosing between them This is an example in which the values of effectiveness and money are not comparable If x is a medicine which is very effective in curing the illness and y and z are two medicines which are both less effective than x, but equal to each other in this respect, although y is less expensive than z, it would be likely that x would be preferred to y, and y to z However, if this is truly a case of non-comparability, then there would be no probability $p > 0$ such that y would be preferred to $<px, (1-p)z>$, since even if p is very small, $<px, (1-p)z>$ would be a probability mixture giving a higher probability of curing the illness than y

In the above example, it seems reasonable to separate the bases of preference for the medicines into two *components*, the first component representing the effectiveness of the medicine, and the second, its cost Likewise, one might assume that the man has two utilities for these alternatives, one utility for the effectiveness in curing the illness, and the other for cost In this case, the man's preferences are represented by a vector with two components In choosing between alternatives, the man chooses the medicine with the highest utility in the first component (representing effectiveness), ignoring the second component entirely unless the two alternatives are equal in the first component, in which case he chooses the alternative with the highest utility in the second component

There is no reason to limit the number of possible components of preference to two, as in the above example, and, in fact, choice situations with larger numbers of components are conceivable In general, the number of non comparable components will determine the number of components in the

utility "vectors" representing the preferences What Hausner has shown is that for any system of preferences which satisfy essentially all of the Consequences of the Bernoullian Utility hypotheses given in Section 4, with the exception of 3 6, it is always possible to represent the preferences by utility vectors' (of unspecified dimension) If the dimension of the vectors (*i e*, the number of their components) is finite, say of order n, then they can be so chosen that the preferences are made according to the rule that an alternative represented by the vector $(x_1, x_2, \ldots, x_n)$ is preferred to an alternative represented by the vector $(y_1, y_2, \ldots, y_n)$ if and only if at the first component at which these vectors differ, say the i^{th} (in which case $x_1 = y_1, x_2 = y_2, \ldots, x_{i-1} = y_{i-1}$), the i^{th} component of x is greater than the i^{th} component of y In the example of the medicines, the preferences would be represented by two components, and (x_1, x_2) would be chosen over (y_1, y_2) if and only if either $x_1 > y_1$ or $x_1 = y_1$ and $x_2 > y_2$ The rule for determining the utility vector representing a probability mixture $<px, (1-p)y>$ of two vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ is analogous to the corresponding rule for ordinary utilities the utility vector of this mixture is simply $(px_1 + (1-p)y_1, \ldots, px_n + (1-p)y_n)$ It can be seen that the ordinary utility theory can be interpreted as simply a special case of this multi dimensional theory, in which the preferences are represented by vectors with just one component

5 3 Theories of Subjective Probability and Utility

We noted in the introduction to this section that there are two ways of dealing with the problem which arises from the fact that subjects may not be aware of the true or "objective probabilities which obtain in decisions which they may face involving risk One way is to restrict application of the theory to situations in which all of the objective probabilities are simple (say ½) and the subjects are aware of them This procedure might seem unduly restrictive, but is quite reasonable to *test* Bernoullian Utility as a descriptive theory in such artificially schematized situations before attempting to construct theories to deal with behavior in more complicated decisions Davidson, Siegel, and Suppes [1957] have followed this course in their experiments testing the Bernoullian Utility hypothesis The second approach is to assume that people are not aware of the objective probabilities, but nevertheless estimate them, and assign 'subjective" probabilities on the basis of which they make their decisions Such an assumption is at the heart of three different theories which are otherwise quite different One of these is due to Ward Edwards [1954b], one to L J Savage [1954], and one to R Duncan Luce [1957] Edwards' theory is a straight-

forward descriptive one, which makes no claim to "rationality" in the sense that it describes the behavior of a rational man Savage gives a set of axioms very much like those described in Section 4, which he argues for as postulates of rational behavior Unfortunately, there is not space here to describe Savage's theory, since it is built on a formalism which is too complicated to be described briefly What we shall do instead is to indicate the central ideas of the theory of subjective probability within a simple framework much like that employed for the "classical" theory, first formulating the theory in terms of the two unobserved functions utility and subjective probability, and then deducing some of its more important behavioral consequences We shall, in the process, derive a behavioral principle closely analogous to the famous "*sure thing principle*" which is the cornerstone of the rational justification of Savage's theory Luce's theory, besides being a theory of subjective probability, includes certain additional features, which demand a separate discussion It will be described in Section 5 5

The alternatives in the subjective probability theory are mixtures in the extended sense described briefly in Section 3 That is, given an event E and its "complement" $\bar{E}$ (the event which happens if E does not), and two sure outcomes x and y, we can form the mixture $<Ex, \bar{E}y>$, which is the alternative of getting x if E happens and y if $\bar{E}$ happens (*i e*, if E does not) To give a specific example, suppose that a man is considering whether to go and visit a friend, but is not sure the friend is at home If he takes the action of going to the man's house, this has two possible outcomes seeing the friend, or not seeing him, and the actual outcome depends on whether the friend is home when he gets there or not If we let E be the event that the friend is at home (and therefore $\bar{E}$ is the event he is away), and x be the outcome of visiting the friend and y be the outcome of making a fruitless trip, then the alternative of going to the friend's house is the mixture $<Ex, \bar{E}y>$

In general, what is assumed in any decision situation is that there is some basic set S of events E The set is assumed to include, besides any event E, its "complement' $\bar{E}$, and for any two events E_1 and E_2 their "union" $E_1 \cup E_2$, which is the ' event" of either E_1 happening or E_2 happening One example would be that the set S consists of all the possible ways a single die can fall We might let the event E_i, $i = 1, \quad , 6$ be the event that the die rolls i (*e g* , E_3 is the event of rolling a three) But, it is assumed that S must contain besides $E_1, \quad , E_6$ all complements $\bar{E}_i$, and all unions $E_i \cup E_j$, unions of unions, etc Thus, S must contain the event $F_2 \cup E_4 \cup E_6$, which represents rolling a 2 or 4 or a 6, *i e* , rolling an even number It is not difficult to show that in this case S must contain 64 separate events Since we shall always be

working with finite sets S, it can be shown that S can always be taken as the set of all subsets of some basis set V, which in the example is the set E_1, E_2 E_3, E_4 F_5 F_6

Now, if we let K_0 be the set of all 'pure or riskless alternatives the set K of mixture alternatives will include besides K_0 all mixtures $<Ex, \bar{E}y>$ where F is in S and x and y are in K_0, mixtures of mixtures, etc (the precise formal definition of K in terms of S and K_0 offers no difficulty, and will be omitted here)

Besides the set K of alternatives, which has now been described, there remain the preference relations P, R, and I and the utility function, which play the same role as in the classical theory, and one new function Π defined over the set S For any event E, $\Pi(E)$ is intuitively interpreted as the subjective or perceived probability of E In Savage's theory, Π is required to have very special properties namely, Π is required to be a *probability measure* over the set S Edwards places weaker restrictions on Π, which admit the possibility that subjective probabilities do not necessarily satisfy even the logical laws of probability (such as the requirement that Π always lie between 0 and 1) It is possible to state the central hypothesis of subjective probability without any specific assumptions about Π, and we shall do so before enquiring into any of the special consequences which follow from such assumptions This hypothesis is simply for all x and y in K and E in S,

$$u(<Ex\ Fy>) = \Pi(E)u(x) + \Pi(E)u(y) \tag{1}$$

Thus formulated, the basic hypothesis is seen to be nothing more than a generalization of the Bernoullian Utility hypothesis If the probabilities $\Pi(E)$ and $\Pi(\bar{E})$ are objective, then $\Pi(\bar{E}) = 1 - \Pi(E)$, and the right side of equation (1) assumes the same form as the analogous equation in the classical theory

In assessing the theory incorporated in equation (1) it is necessary to derive its directly testable consequences These are the ones which are stated solely in terms of the observable relations P, R and I between the alternatives, and do not involve the unobservable functions u and Π (recall that in the subjective theory $\Pi(E)$ is a *subjective* magnitude not accessible to direct inspection) Even without any further assumptions about the form of Π(such as Savage makes) certain consequences follow from equation (1) which are immediately testable One trivial one is that preferences must be transitive Unfortunately, indifferences must also be transitive and so this theory suffers from at least some of the defects of the classical theory An

other not quite so obvious though expected consequence is that if $\langle Ex, \bar{E}y \rangle$ is preferred to $\langle Ez, \bar{E}y \rangle$, then $\langle Ex, \bar{E}w \rangle$ is preferred to $\langle Ez, \bar{E}w \rangle$.

The foregoing consequences were cited to show that even without any special hypotheses about Π, the hypothesis formulated in equation (1) has empirical content, and is theoretically refutable. It would be interesting if someone were to derive a set of *sufficient* empirical consequences of equation (1) which, if they were satisfied, would imply that there were functions Π and u satisfying the equation. To the author's knowledge this has not yet been done. For present purposes, however, the theory becomes much more interesting if Π is assumed to have a special form. As noted above, in Savage's theory Π is a probability measure (*). The conditions Π must satisfy in order to be a probability measure can be stated as follows: for all E_1 and E_2 in S,

$$0 \leq \Pi(E_1) \leq 1, \tag{2}$$

$$\text{if } E_1 \cap E_2 = 0, \text{ then } \Pi(E_1 \cup E_2) = \Pi(E_1) + \Pi(E_2), \tag{3}$$

$$\Pi(V) = 1 \tag{4}$$

(In the above equations $E_1 \cap E_2$ is the "intersection" of the events E_1 and E_2, the event of both E_1 and E_2 happening simultaneously, 0 is the null event, and $E_1 \cap E_2 = 0$ simply says that E_1 and E_2 are logically incompatible, i.e., if one happens, the other does not. V is the universal event, the one which is sure to happen.) One very important consequence of equations (2), (3), and (4), is that: for all E in S,

$$\Pi(E) + \Pi(\bar{E}) = 1 \tag{5}$$

One immediate consequence of equations (1) — (5) is a simple kind of sure thing principle as follows:

$$\text{if } xRz \text{ and } yRz, \text{ then } \langle Ex, \bar{E}y \rangle \; Rz \tag{6}$$

This says that if x and y are both at least as good as z, then any mixture of them is at least as good as z. It is obvious why such a consequence is called a "sure-thing principle", the alternative $\langle Ex, \bar{E}y \rangle$ is a sure-thing by comparison to alternative z, since no matter which outcome x or y results from, taking alternative $\langle Ex, \bar{E}y \rangle$, this is at least as good as z.

The important problem with this theory, as with other theories involving unobserved functions, is to derive a set of behavioral consequences which

(*) Savage formulates his theory directly in terms of the behavioristic consequences and then derives the existence of the functions Π and u, rather than, as we do here, assume the existence of these functions, and then derive their empirical consequences.

are equivalent to the original assumptions One way to do this is to arrive at definitions of the unobserved functions Π and u in terms of observations, since with Π and u observationally defined, the fundamental hypotheses become observationally testable Unfortunately, it is not possible to give the required definitions in this case without making further assumptions It is also impossible to formulate any simple set of observable consequences which are logically equivalent to the original assumptions Hence, the best that can be hoped for is to give sets of behavioral consequences which are intuitively significant and test them It turns out that for most of these consequences there are strong arguments as to why a rational person should act in accordance with them (the simple sure thing principle mentioned above is an example), and therefore this theory of subjective probability can be defended as a theory of rationality

5 4 Theories of Imperfect Discrimination

The problem raised by the fact that individuals are not in general able to compare utility differences with perfect accuracy, and hence their indifference judgments are likely to be intransitive, is one which Bernoullian Utility theory has in common with any other theory of utility which makes the fundamental assumption that x is indifferent to y if and only if $u(x) = u(y)$ As a result, most of the theories which have been advanced to take into account imperfect discrimination of utilities do so within the framework of utility theory in general rather than within the more restricted scope of Bernoullian Utility In fact the theory of imperfect discrimination has its origin outside utility theory completely, and inside the domain of psychophysical measurement, particularly in the theory of sensations of tone The detailed discussion of these theories therefore carries us considerably outside the scope of this survey, and we shall limit ourselves to a few sketchy remarks

Traditionally, there are two ways of approaching the problem of imperfect discrimination The first of these postulates that rather than indifference being equivalent to equality of utility value, this relation should hold if the two alternatives compared are less than *just noticeably different* in utility value Specifically, it is possible for two alternatives to differ in utility value and yet be regarded as indifferent because they are not noticeably different in utility With respect to utility theoretic applications this approach has been followed by Luce [1956] and Gerlach [1957] A second approach to the problem is *stochastic* it assumes that preference judgments are subject to random fluctuations, and that intransitivities of preference or

indifference are not due to imperfect discrimination but rather to random fluctuations in the underlying preference scale This type of assumption has also a long history in the literature of psychological scaling, and has been specifically applied to utility theory by Davidson and Suppes [1957] and by Luce [1957], whose theory we will examine below.

One very simple formulation of the theory of imperfect discrimination based on just noticeable differences makes the following assumptions there is some $\Delta > 0$ such that for all x and y in K,

$$xPy \text{ if and only if } u(x) > u(y) + \Delta, \tag{1}$$

$$xIy \text{ if and only if } | u(x) - u(y) | \leq \Delta \tag{2}$$

One immediate consequence of these two assumptions is that the two relations P and I compose what Luce [1956] has called a "semi-order" Scott and Suppes [1958] have given the following axioms for semi-orders for all $x, y, z,$ and w in K,

$$\text{not } xPx, \tag{3}$$

$$\text{if } xPy \text{ and } zPw, \text{ then either } xPw \text{ or } zPy, \tag{4}$$

$$\text{if } xPy \text{ and } zPx, \text{ then either } uPy \text{ or } zPu \tag{5}$$

Indifference is defined by the condition

$$xIy \text{ if and only if not } xPy \text{ and not } yPx \tag{6}$$

Scott and Suppes proved that for finite sets of alternatives the semi-order axioms given above exhaust the empirical significance of assumptions (1) and (2) It might be thought that an essential restriction is imposed by the assumption that the number Δ, which defines the interval of indifference, is constant and does not depend on either $u(x)$ or $u(y)$ Intuitively one might think that the accuracy of discrimination might vary over the utility scale, possibly becoming less for larger and larger utility values, and therefore admit the possibility that Δ should be variable It turns out, though, that, at least for finite systems, the only behavioral consequences of this assumption are just the semi-order axioms, and hence the theory with the variable j n d (just noticeable difference Δ) is equivalent as far as the behavior it implies, with the theory based on a constant j n d If, however, further conditions are imposed on the utility functions besides (1) and (2), then it can be shown that the assumptions of constant and variable j n d 's are different. Some of these problems have been investigated by Gerlach [1957]

The stochastic approach to the phenomenon of imperfect discrimination involves a somewhat more radical reformulation of the underlying conceptual apparatus than does the j n d formulation Here the relations P and I are replaced by probabilities p and i, where $p(x, y)$ is taken to be the *probability* that on any given occasion a person will choose x in preference to y, and $i(x, y)$ is the probability that they will be regarded as indifferent If on any given occasion the person must do exactly one of the following three things — choose x over y, declare himself indifferent between x and y, or choose y over x — then it follows that the three probabilities $p(x, y)$, $i(x, y)$, and $p(y, x)$ must add up to 1

$$p(x,y) + i(x,y) + p(y,x) = 1 \tag{7}$$

In view of the fact that indifference between y and x is logically equivalent to indifference between x and y, we also have

$$i(x,y) = i(y,x) \tag{8}$$

In most stochastic theories of choice it is assumed that the subject cannot declare himself indifferent between two alternatives — he is forced to choose one or the other In this case $i(x,y) = 0$, and equation (7) reduces to

$$p(x,y) + p(y,x) = 0 \tag{9}$$

It is usually assumed that the probabilities are connected with the 'true' preferences around which the actual choices fluctuate randomly in some such way as the following

for all x and y in K,

$$xPy \text{ if and only if } p(x,y) > \tfrac{1}{2}, \tag{10}$$

$$xIy \text{ if and only if } p(x,y) = \tfrac{1}{2} \tag{11}$$

Now, if the preferences and indifferences satisfying conditions (10) and (11) are assumed to satisfy some theory of utility, such as the Bernoullian, then the observable consequences of that theory become statements about probabilities, and can be tested by using the observed relative frequencies of choice as estimates of these probabilities

Before passing on to consider Luce's theory, it is well to note that the stochastic formulation of the theory of preference gets around the problem of intransitivity of indifference by getting into what is in many ways an even thornier problem namely, the problem of estimating probabilities from finite relative frequencies If the number $p(x\ y)$ is obtained as the re-

lative frequency with which x is chosen over y in a finite number of observations, then almost certainly neither preference nor indifference as defined by conditions (10) and (11) will be transitive, *and it would appear that we* are in a worse difficulty than before. The traditional statistical solution is to regard these numbers as estimates of the true probabilities and subject to sampling error. The question then arises as to when a given set of observed relative frequencies $p(x,y)$ which does not perfectly satisfy a theory does not do so because of sampling error, or because the true probabilities actually do not satisfy the hypotheses of the theory. The answer to this question would take us into the statistical theory of testing hypotheses, which will not be embarked upon here. Suffice it to say that the advantage of avoiding intransitivity of indifference is paid for in the extreme difficulty of actually applying or testing the stochastic theory with which it is replaced.

5.5 Luce's Theory of Discrimination Structures.

We conclude the discussion of modifications of the classical theory with a theory which combines both subjective probabilities and a stochastic treatment of imperfect discrimination. This is Luce's theory of *discrimination structures* [1957]. This model also has one novel feature of great interest: it assumes not only imperfect discrimination of utilities, but also imperfect discrimination of subjective probabilities, and advances an interesting hypothesis relating the two.

A real valued function p over $K \times K$ (the "cartesian product" of K with itself) is said to be a *discrimination structure* if it satisfies the following three conditions:

$$p(x,y) \geq 0 \text{ for all } x,y \text{ in } K, \tag{1}$$

$$p(x,y) - p(y,x) \leq 1 \text{ for all } x,y \text{ in } K, \tag{2}$$

$$\text{for some } a \text{ and } b \text{ in } K,\ p(a,b) = p(b,a). \tag{3}$$

Intuitively, the number $p(x,y)$ may be thought of as the probability that x will be chosen over y on any given occasion. Conditions (1) and (2) simply state some of the properties which p must have given this interpretation. Condition (3) has the function of requiring that not all alternatives be indifferent.

So far the theory differs in no way from the stochastic theory of imperfect discrimination briefly described above. Two theories of discrimination structures for mixture alternatives have been developed by Luce, one of them applying to what he calls "*weak mixture spaces*," closely analogous to the non-

numerical mixtures described briefly in Section 3, and the other applied to probability mixtures in the ordinary sense Since the parallel with the classical theory is more obvious in the case of discrimination structures over ordinary mixture spaces we shall confine our attention to the second theory (actually, Luce considers what he calls "*risk spaces*" which do not satisfy all the axioms of mixture spaces) The results he finds for risk spaces, however, apply *a fortiori* to mixture spaces, and we shall assume for the remainder of this section that K is an ordinary mixture space In stating the fundamental axiom of the theory of discrimination structures on mixture spaces, it is necessary to introduce a second discrimination function q, analogous to p, but representing discrimination of relative likelihoods of pairs of events If α and β are the probabilities of two events, say E and E (it is necessary here to use the Greek letters, since the letters 'p' and 'q" are now being used to denote the two discrimination functions), then $q(\alpha,\beta)$ is interpreted intuitively to be the probability that event E is judged by the subject as more likely to occur than event E' It is important to note that the above description gives only an intuitive interpretation of q the fundamental assumption of the theory of discrimination structures on mixture spaces does not presuppose any objective interpretation for q, but only that a function q with a certain property *exist* The central assumption is that a discrimination structure p on a mixture space K is *decomposable*, as defined below

Definition (Decomposable Discrimination Structures). A discrimination structure p on a mixture space K is said to be decomposable if there exists a real-valued function q, which is called the *core* of p, such that for all α $\beta \epsilon [0,1]$,

$$q(\alpha,\beta) \geq 0,$$

$$q(\alpha,\beta) + q(\beta,\alpha) \leq 1,$$

for all $x,y \epsilon K$,

$$p(<\alpha x, (1-\alpha)y>, <\beta x, (1-\beta)y>) = p(x,y)q(\alpha\ \beta) + p(y,x)q(\beta,\alpha)$$

The last axiom in the definition of a decomposable discrimination structure is of particular importance, since it stipulates the relationship between the functions p and q This axiom is justified by the following intuitive argument There are intuitively two possible cases in which the mixture $<\alpha\ x,(1-\alpha)y>$ will be preferred to the mixture $<\beta x, (1-\beta)y>$ (1) the case in which x is preferred to y, and the event with probability α is regarded as more likely to occur than the event with probability β, and (2) the case in which y is preferred to x, and the event with probability β is regarded as more likely to occur than the event with probability α If it is assumed that

the discriminations of preference are statistically independent of the discriminations of relative likelihood, then the probability that the subject will both judge x to be preferred to y, and judge the event with probability α to be more likely than the event with probability β will simply be the product $p(x,y)q(\alpha\ \beta)$ Similarly the probability that he will judge y to be preferred to x, and β to be more likely to occur than α will be $p(y,x)q(\beta,\alpha)$ According to the foregoing argument, then the total probability that he should prefer $<\alpha x, (1-\alpha)y>$ to $<\beta x, (1-\beta)y>$ should be the sum of the two probabilities $p(x,y)q(\alpha\ \beta) + p(y,x)q(\beta,\alpha)$ This is just what condition (iii) of the definition — the so-called "*decomposability* axiom — states

It is to be observed that in the above argument it is essential to regard the function $q(\alpha\ \beta)$ as representing the probability that the subject will regard an *event* with probability α as more likely to occur than an *event* with probability β, but he is not presumed to know the two probabilities α and β while discriminating between their relative likelihoods If the subject did know the probabilities α and β before deciding which of the two events was the more likely to occur, he would certainly choose the event for which the probability was the greater *i e*, it would be the case that $q(\alpha,\beta) = 1$ if α were greater than β, and $q(\alpha\ \beta) = 0$ if β were greater than α The problem of giving a satisfactory interpretation to the numbers α and β is now seen to be a thorny one One solution proposed by Luce is not to begin with numbers at all, but with events instead, and, using the assumptions of the theory, to determine numbers corresponding to the events in such a way that the decomposability axiom and the mixture space axioms are satisfied A discussion of this proposal would take us beyond the scope of this survey, and we may once again refer the interested reader to Luce's paper

As in the case of the axioms for the classical theory of Bernoullian utility, one of the important questions which can be raised about the theory of discrimination structures on mixture spaces is under what conditions does there exist a Bernoullian Utility function (what Luce calls a "linear" utility function) for such systems? Luce shows that every decomposable discrimination structure which satisfies an Archimedian" axiom very similar to the Archimedian axiom formulated in Corollary 3 6 (to Consequence 3 in Section 4) has a Bernoullian Utility function What is perhaps more interesting is the fact that these Bernoullian Utility functions are also shown to be *sensation scales*, since $p(x,y)$ is shown to be a non-decreasing function of the utility difference $u(x) - u(y)$

6 THE DECISION PROBLEM

6 1 Description of the Decision Problem and its Relation to Utility Theory.

The *decision problem* is the problem of choosing in what is in some sense the 'best' way among a number of alternative courses of action, and a *decision theory* is any theory which provides a systematic solution to a decision problem Clearly, utility theory and Bernoullian Utility theory in particular are decision theories in the sense described above, since they provide decision rules prescribing which action or alternative to take in situations of certain kinds (namely those involving *risk*) There are, however, certain types of decisions which cannot be treated within the scope of Bernoullian Utility theory, because either they involve factors determining the outcome of the action which are not simple risks, or because the factors even if they are risks, are such that their objective probabilities are not known We shall examine briefly in this chapter some decision theories which provide solutions to the decision problem in non risk situations There are two reasons why a knowledge of other decision theories is important to the understanding of Bernoullian Utility theory First, many of the decision rules of the other theories are different from those of Bernoullian Utility, and therefore the axioms (or some of them) of Bernoullian Utility do not hold in the situations in which the other theories do Hence these other decision theories show in a sense some of the limitations of Bernoullian Utility Secondly, Bernoullian Utility is often incorporated into the foundation of other decision theories, and therefore these decision theories are *applications* of Bernoullian Utility

The general decision situation involves the following factors (1) a set of possible *actions* which an individual may take, (2) a set of factors or *states of nature* which are beyond the individual's control, but, together with the actions, determine (3) the *outcome* Very roughly speaking, the above three factors — actions, states of nature and outcomes — correspond in Bernoullian Utility theory to the *mixture alternatives*, the *events* which determine the probabilities in the mixtures, and the *pure alternatives* which are the final outcomes Beyond the physical situation described above, there are *preferences* In the general decision case there will be two systems of preferences one among the set of possible outcomes, and the other among the alternative actions The decision problem can be formulated as asking given the preferences among the outcomes, what should be the preferences among the actions? In Bernoullian Utility theory this problem is solved in a very simple

way the utility of the action (*i e*, the measure of the strength of preference for it) must be the *expected value* of the utilities of its outcomes when these utilities are each multipled by the probability with which that outcome will occur

Decision theory as formulated above includes three important disciplines within its scope These are (1) utility theory, including Bernoullian Utility theory, (2) statistical decision theory, and (3) the theory of games Statistical decision theory is, roughly, the theory of decision-making based on statistical data It is concerned with the sort of problem a statistician faces when he has to decide on the basis of sample data whether or not to accept or reject a certain lot of goods which may be defective The theory of games is, as its name implies, the theory of decisions made in "games," or, more generally, competitive situations, in which the unknown factors determining the outcome in any situation include the acts of rational "opponents" Historically, the theory of games is very important to Bernoullian Utility theory, since it was as an adjunct to their theory of games that von Neumann and Morgenstern created the modern theory of Bernoullian Utility (*The Theory of Games and Economic Behavior*, 1947) We shall discuss briefly the conceptual bases and formalisms of statistical decision theory and game theory later in this chapter

Every decision theory can, like Bernoullian Utility theory, be interpreted either as a descriptive or as a normative theory, and the interpretation given in any instance will depend on the use to which the theory is to be put In what follows, however, the decision theories will be discussed primarily from the point of view of rationality The reason for this is simply that their assumptions are more plausible when viewed as canons of rationality rather than as descriptions of actual behavior This approach is partly justified by the fact that until now there has—with one exception (an experiment by Suppes and Atkinson [1958]) — been no attempt to apply any decision theory other than utility theory to the explanation of actual behavior in decision situations We shall examine two descriptive applications of Bernoullian Utility in Section 7

The present discussion of decision theory is necessarily brief, and the reader desiring a more extended and thorough analysis should consult other references Blackwell and Girshick's *Theory of Games and Statistical Decisions* [1954] is an excellent introduction to games and statistics as decision theories Savage's *Foundations of Statistics* [1954] also deals with statistics as a decision problem, and is particularly thorough in its analysis of the various decision principles which have been proposed as solutions of the decision problem

6 2 Formalization of the Decision Problem (*)

6 2 1 Decision Situations — Matrix Representation

A *decision situation* corresponds to what can be thought of as the purely physical aspect of a decision problem A decision situation is comprised of four elements (1) a set A of possible *acts*, (2) a set S of possible *states of nature*, (3) a set O of possible *outcomes*, and (4) a function M such that if a is a member of A (i e , a is an act), and s is a member of S (s is a state of nature), then $M(a,s)$ is the member of O (i e , the outcome) which results if act a is chosen and s is the actual state of nature Formally, A, S, and O are sets, and M is a function mapping $A \times S$ into O A is sometimes called the set of "*strategies*," this term being derived from game theory, in which the action of the individual consists in his choosing a strategy of play The function M is sometimes termed the "*outcome function*," because it determines what the person making the decision gets as a result of taking whatever act he chooses Note particularly that there is nothing in the decision situation representing the individual's preferences, either among outcomes or among acts

Decision situations in which there are only a finite number of possible acts and a finite number of possible states of nature are often represented by matrices If there are n members of the set A, say $A = \{a_1, \quad , a_n\}$, and m members of the set S, say $S = \{s_1, \quad , s_m\}$, then the decision situation is represented by a matrix with n rows and m columns, in which the entry in the i^{th} row and the j^{th} column is $M(a_i, s_j)$, i e , the entry is the outcome which results from action a_i and state of nature s_j An example will illustrate this method of representation

Suppose that a man had the opportunity either to bet \$10 on Eisenhower or on Stevenson in the 1956 presidential election, or else not to bet In this case there are three possible actions a_1 = bet \$10 on Eisenhower,

	s_1 = Eisenhower wins	s_2 = Stevenson wins
a_1 = bet \$10 on Eisenhower	$M(a_1\ s_1)$ = win \$10	$M(a_1\ s_2)$ = lose \$10
a_2 = bet \$10 on Stevenson	$M(a_2\ s_1)$ = lose \$10	$M(a_2\ s_2)$ = win \$10
a_3 = don t bet	$M(a_3\ s_1)$ = break even	$M(a_3\ s_2)$ = break even

(*) The notation employed here for decision situations preferences and utilities differs in some respects from other systems of notation although there is as yet no perfectly uniform system of notation for these systems

a_2 = bet \$10 on Stevenson, a_3 = not bet There are only two relevant "states of nature" in this example s_1 = Eisenhower wins election, and s_2 = Stevenson wins election Finally, there are three possible outcomes to be considered win \$10, lose \$10, or come out even This situation is represented in the matrix on the preceding page

Normally, the matrix would be simplified as follows

	s_1	s_2
a_1	\$10	—\$10
a_2	—\$10	\$10
a_3	\$0	\$0

There is a close connection between decision situations and what we have called "weak mixtures," discussed in Sections 3 and 5 The situation in the above example could equally well have been represented as a decision among three weak mixture alternatives, in which the act a_1 would correspond to the alternative $<E(\$10), \bar{E}(-\$10)>$ where E is the event of Eisenhower's winning the election Similarly a_2 would correspond to the mixture $<E(-\$10), \bar{E}(\$10)>$, and a_3 would correspond to $<E(\$0), \bar{E}(\$0)>$ In general there is a straightforward way of translating decision situations into systems of weak mixture alternatives and vice versa, though we shall not define this connection formally here

Before going on to consider the individual's preferences, there are two points of particular importance to notice about decision situations and their corresponding matrix representations First, the set of states of nature are here construed to include only the factors in the situation which are not under the individual's control, but which in some way determine the outcome of his act In the example, the relevant states of nature were just Eisenhower's winning or Stevenson's winning, since once the man made his bet the only thing that determined its outcome was which of these two men won the election There is no harm (in most cases) in including irrelevant factors — such as the winner of the 1956 World Series — among the states of nature, but these have no bearing on the solution of the decision problem The second thing which it is important to keep in mind is the fact that the outcomes $M(a_i,s_j)$ which are entered in the spaces in the matrix are the actual physical

results For example $M(a_1\ s_1)$ is the physical event of the bettor's receiving a ten dollar bill or its equivalent Specifically, no assumption is yet made about the valuation which the individual attaches to these outcomes, and different individuals with different sets of values may be confronted with the same decision situation The next step is to incorporate formally the element of preference in the decision situation

6 2.2 Preferences.

As was indicated above, there are, at least conceptually, two sets of preferences involved in a decision problem preferences among outcomes and preferences among acts Logically, the system of preference relations over the sets A and O (acts and outcomes) must be kept distinct However, since A and O have no members in common (*i e*, no act is at the same time an outcome and vice versa), no confusion will result if we continue to use the symbols P, I, and R for preference, indifference, and preference or indifference relations of both A and I The three relations P, I, and R will again be assumed to have the properties and inter relations assumed in Section 3 (see page 174)

6 2 3 Utility.

In analyzing some decision problems the concept of utility is introduced at the outset, along with the preference relations As in the case of these relations, utility is logically represented by two functions, one over the set A and the other over the set O However, we shall continue to use letter u to denote both utility functions As before, it will be assumed to satisfy at least the ordinal assumption for all x and y in A, $u(x) > u(y)$ if and only if xPy, $u(x) = u(y)$ if and only if xIy, and $u(x) \geq u(y)$ if and only if xRy The same connection between utility and the preference relations holds for any x and y in O Later it will also be required that u satisfies the expected utility assumption, but so far no probability mixtures either of acts or outcomes have been introduced

The introduction of utilities for the outcomes allows an important simplification in the matrix representation of decision situations Most of the essential information in the outcome matrix is retained if this matrix is replaced by an associated '*payoff matrix*, and the outcome function M is replaced by a *payoff function* U which for each act a and state of nature s determines the utility value of the outcome $M(a,s)$ Formally the payoff function is defined as follows for all a in A and s in S,

$$U(a,s) = u(M(a,s))$$

The payoff matrix is like the original outcome matrix except that each of the outcomes in its spaces is replaced by its utility. The justification for replacing the physical outcome $M(a,s)$ by its utility $U(a,s)$ is that the solution to the decision problem — which is that of determining a set of preferences among the acts, given the preferences among the outcomes — depends only on the utilities of the outcomes, and on the state of nature, and not on any other feature of the outcome.

In general, if A and S are finite sets, $A = \{a_1, \ldots, a_n\}$ and $S = \{s_1, \ldots, s_m\}$, then the decision problem will be represented by the following payoff matrix

States of Nature

Acts	s_1	s_2	...	s_j	...	s_m
a_1	$U(a_1, s_1)$	$U(a_1, s_2)$		$U(a_1, s_j)$		$U(a_1, s_m)$
a_2	$U(a_2, s_1)$	$U(a_2, s_2)$		$U(a_2, s_j)$		$U(a_2, s_m)$
⋮						
a_i	$U(a_i, s_1)$	$U(a_i, s_2)$		$U(a_i, s_j)$		$U(a_i, s_m)$
⋮						
a_n	$U(a_n, s_1)$	$U(a_n, s_2)$		$U(a_n, s_j)$		$U(a_n, s_m)$

The decision problem can now be stated formally: given the above payoff matrix, determine the preference relations P, I, and R, and the utility function u for the set A of acts. Before going on to describe some of the solutions for special cases of the decision problem, it is necessary to describe certain modifications of the above basic decision situation, which arise in statistical decisions, and in games.

6.3 Mixed Strategies.

For some purposes, particularly in the theory of games, it is desirable to extend the set A of possible actions in a decision situation by including what are called "*mixed strategies*" or "random strategies." A mixed strategy is, formally, a probability mixture of actions. The intuitive idea of a mixed strategy is simply that the individual, instead of choosing an act outright in

a decision situation, may decide to allow the action he will take to be chosen by a random device For example, in the case of the man who must decide whether to bet on Eisenhower or on Stevenson, or not to bet on the outcome of the 1956 presidential election, he could have decided to flip a coin, and bet on Eisenhower if it came down heads and bet on Stevenson if it came down tails This is a mixed strategy in which there is a 50 % probability that he bets on Eisenhower, a 50 % probability that he bets on Stevenson, and zero probability that he does not bet

Formally the act of choosing a mixed strategy is equivalent to choosing a probability distribution over the set A If σ is a probability distribution over A, then for each act a in A, the probability of its being chosen is simply $\sigma(a)$ The set of all mixed strategies over A, called the '*mixed strategy* space,' will be denoted 'Σ' In case A is finite, Σ will just be the set of all probability distributions over A

It is clear that the sets Σ, S, *and* O *determine a new decision situation* derived from the basic decision situation which is determined by A, S, O, and the outcome function M That is, the act of choosing a mixed strategy over A, *together with the actual state of nature will determine the outcome* of the decision in a sense It remains, however, to define the outcome function for this derived decision situation Suppose that the individual decides to follow the mixed strategy represented by the distribution *function* σ *over* A, and s is the state of nature Then with probability $\sigma(a)$ he will take action a, and, *since the state of nature is* s, the outcome will be $M(a,s)$ Hence outcome due to taking mixed strategy σ and state of nature s *is a mixture of* outcomes, in which the individual gets outcome $M(a,s)$ with probability $\sigma(s)$

If the outcomes $M(a,s)$ *are replaced by their utilities* $U(a,s)$, then the payoffs due to taking mixed strategies are easily defined *If the utilities obey* the expected utility hypothesis, then the utility of the mixture outcome from mixed strategy σ and state s *must simply* be the expected value of the utilities of the 'pure outcomes Hence (in case A is finite), the payoff $U(\sigma,s)$ *from mixed strategy* σ *and state* s *is simply*

$$U(\sigma,s) = \sum_{a\in A}\sigma(a)\,U(a,s)$$

6 4 Statistical Decisions

The basic decision formalism, with either the outcome *function* M *or the* payoff *function* U, *is* quite a flexible mold, and, depending on the interpretations given the sets A and S (or Σ), can be applied to a variety of problems Here and later in this chapter we will show how both statistical decisions and

games can be incorporated into this model if the appropriate interpretations of A and S are made In each case we shall only show how the decision *problems* in games and statistical decisions are formulated, without indicating any of the proposed solutions Solutions to various cases of the general decision problem, including games and statistical decisions, are discussed in Section 6 6

The essential characteristic of a statistical decision is that it involves taking an action on the basis of statistical information This statistical information generally gives some indication as to the actual "state of nature" which prevails, although it does not as a rule determine what the true state of nature is with certainty A typical statistical decision is the following Suppose that a manufacturer of electrical equipment has received a shipment of 1000 fuses, but he fears that there is a high proportion of defective fuses in the lot If the proportion is too high, it would be more profitable to him to return the fuses to the shipper, rather than put them into the equipment he manufactures He decides to take a sample of 20 of these 1000 fuses, test them to determine how many in the sample are defective, and make his decision as to whether to return the 1000 or not on the basis of result of his test In this instance, the manufacturer's action or "strategy" consists in taking a sample of 20, and deciding on the basis of this whether or not to send the lot back A complete strategy for him consists in his deciding in advance which final action to take for each possible outcome of the observations for instance, he might decide in advance that if he finds more than three defective in the twenty he will return the shipment, and otherwise keep it

A strategy like that described above, in which the manufacturer decides in advance that he will examine all twenty fuses willy-nilly, even though he might find the first four defective and hence not need to examine the remaining 16, is called a "single experiment strategy " In general, single experiment strategies are those in which the operations in the experiment are prescribed in advance (in the example, the operations consisted of putting each of the twenty fuses in a fuse tester) The manufacturer could vary his procedure by deciding to test fuses one at a time until he had found four defective fuses, or until he had tested twenty fuses, and if he found four defective he would return the lot Such a strategy is called a 'sequential strategy" or a "sequential experiment' , sequential strategies are characterized by the fact that the operations are not specified in advance, but in general depend on what observations are made In general, it makes no essential difference whether a single-experiment strategy or a sequential strategy is used

unless there is a '*cost function*,' which sets a cost on the operations involved in the sampling In such a case, it will clearly be to the advantage of the individual making the decision to use a sequential strategy In this section, however, we shall confine our attention to single experiment strategies without a cost function, and show how decisions involving these are incorporated within the framework of general decision theory

As noted above, an "act" in a statistical decision situation consists in choosing a strategy which tells the person what to choose for each of the possible outcomes of the 'experiment' There are three elements involved here (1) a set Z of possible outcomes of the experiment, (2) a set K of choices which the person may make after observing the results of the experiment, and (3) a set A of 'strategies' each of which is a rule which directs the person which final choice k in K to take for each possible observation z in Z Formally, A is a set of functions which map Z into K, i e, for each strategy a in A and observation z in Z, $a(z)$ is the choice in K which the individual who chooses strategy a makes if z is observed The set Z is called the 'sample space,' K is called the set of 'terminal actions' and A is called the 'strategy space' The strategy space in a statistical decision situation is simply the set of acts for that situation The only difference between the statistical decision situation and the general decision situation as far as concerns the set of acts is that an 'act' in a statistical decision situation is not the final or 'terminal' action, but a complex strategy

In the example of the fuses, the sample space, set of terminal actions, and strategy space are as follows If the sampling procedure consists in testing 20 fuses, then there are 21 possible results of this test, since the important information is just the number found defective in the 20, and there can be any number from 0 to 20 defective out of 20 We can define z_n to be the result in which n fuses are found defective There are just two possible terminal acts in this situation to accept the shipment, or to reject it Hence, K has just two members, and we may denote these k_1 (accept) and k_2 (reject) Finally, the strategy space A consists of all possible rules which for each $n = 0, 1, \ldots, 20$ determine which of actions k_1 or k_2 the manufacturer is to take in case z_n (n fuses defective out of 20) is observed Formally, A is just the set of functions mapping Z into K One such strategy is that which stipulates that the lot is to be rejected if four or more defectives are found This corresponds to the function a such that $a(z_n) = k_1$ if $n = 0, 1, 2, 3$, and $a(z_n) = k_2$ if $3 < n \leq 20$

To complete the discussion of the statistical decision situation it is necessary to describe the states of nature, the outcomes and the outcome

and payoff functions It is assumed that the states of nature are of an arbitrary nature, as before, *i e*, S is an arbitrary set However, for any given state of nature s, it is stipulated that there is a calculable probability that any sample z in Z will be observed Therefore, each state s in S is associated with a probability distribution ϕ_s over the set Z, such that if s is the true state of nature, then $\phi_s(z)$ is the probability that the result of the sampling operation will be z Returning to the example of the fuses, the states of nature for the purposes of this problem can be taken to be just the total number of defective fuses out of the lot of 1000 Hence, there are 1001 members of S, and we can define s_m, $m = 0,$, 1000 to be the state in which there are m fuses defective Given that there are m defective fuses, then there is a definite probability that in a sample of 20 fuses drawn at random from the lot there will be n defective This is the probability $\phi_{s_m}(z_n)$, *i e*, the probability that if state s_m holds, then sample z_m will be observed

To define the outcome and payoff functions for the statistical decision situation it is necessary to assume that for each terminal act k and state s there is correlated an outcome $M(k,s)$, and a utility $U(k,s)$ Thus, the sets K and S and the function M constitute a basic decision situation from which the statistical decision situation is derived, much as the decision situation with mixed strategies is derived from the fundamental situation in which there are no random choices The outcomes in the example of the electrical fuses were not specified, but presumably they would be monetary gains or losses, and the payoffs would be the corresponding utilities The outcome of, say, $M(k_1, s_{25})$ is the financial result of taking action k_1 (accepting the lot) in state of nature s_{25} (in which there are 25 defective fuses in the total 1000)

To determine the outcome and payoff functions for the strategy space A and set S of states of nature, we proceed as follows Suppose that a is the strategy chosen, and s is the state of nature Then for each z in Z, $a(z)$ is the terminal action the individual will take if z is the result of the sampling operation, in which case the outcome will be $M(a(z),s)$ and the payoff will be $U(a(z),s)$ Given that s is the state of nature, however, the probability that z will be observed is $\phi_s(z)$ Hence, the outcome $M(a,s)$ resulting from strategy a and state s is a probability mixture of the outcomes $M(a(z),s)$, in which outcome $M(a(z),s)$ occurs with probability $\phi_s(z)$ Again, assuming that the utilities satisfy the expected utility hypothesis, the payoff $U(a,s)$ is simply the expected value of the utilities $U(a(z),s)$ for all a in A and s in S,

$$U(a,s) = \sum_{z\varepsilon Z} \phi_s(z)\, U(a(z),s)$$

It is interesting to note the similarity between the statistical decision situations and the mixed strategies discussed above. Both begin with a basic decision situation, defined by a set of final or "terminal" acts, a set of states of nature, and a payoff function. Then each "extends" the set of strategies in such a way that the individual no longer chooses the terminal act directly, but instead chooses a "mechanism" which will determine the terminal act he will take. In the case of the mixed strategies the mechanism in question was a random device which selected the action, and in the case of the statistical decisions, the mechanism was a decision rule specifying which terminal act to take for each outcome of the sampling process. The set of acts in the extended decision situations then consists of the possible mechanisms which the individual can use to select the terminal action for him: in one case he must choose which random device to use, in the other he must select from among the available strategies. Finally, there is a close similarity between the equations which relate the payoffs in the mixed strategies and in the statistical strategies to the payoffs in the basic decision situations from which they are derived. In fact, the mixed strategy might be regarded as a special case of the statistical decision in which probabilities $\phi_i(z)$ of the samples are independent of the states of nature.

6.5 Games.

6.5.1 The Formalism of Games.

Games in the most general sense are decision situations in which the outcome depends on the actions taken by more than one self interested individual or "player." As such, games could be construed as including almost all social situations, from parlor games to athletic contests to economic transactions. Game-theoretic analysis, however, is most naturally applied to rather rigidly formalized situations like parlor games in which the sets of possible actions are clearly defined by the "rules." Chess, poker, and bridge are all good examples of games well fitted for analysis in game theory because the possible acts or moves at each point are precisely defined by the rules of the game. Athletic contests do not fit the pattern as well because, even though the rules may be fairly clear, the possible actions of the players are limited by physical factors which are hard to define. Similarly, economic situations do not fit so well into the scheme of game theory, in this case because the rules (defined by legislation) governing the transaction are not definite. Attempts have been made to apply game theory to economics and to physical combat situations, but they all involve certain rather artificially simplifying assumptions about the actions which are available to the partici-

pants In this brief discussion, we shall confine our attention to the more clearly defined game situations such as are exemplified in parlor games

An action by a participant in a game consists in his choosing a *strategy*, much as in statistical decision situations, which tells him what "move" to make under all conceivable circumstances which might arise in the course of the play This conception of an act is necessary if games are to be incorporated within the framework of decision theory, because that theory applies only to single acts, and not to series of acts, such as the sequence of moves in playing a game of chess If the free choices which the individual has in the playing of a game are reduced to the choice of a single over all strategy, then each player has only one "action," since once he has chosen his strategy, he can leave the actual playing of the game to an agent whose function it would be to follow mechanically the instructions provided by the strategy at each move As a theory of rationality, the assumption that the player makes a single decision in choosing an over all strategy is in many ways acceptable, since presumably the perfectly rational player in taking any action would survey all of its possible consequences before making his decision But, except in a very limited range of simple games, the assumption that an individual picks an over-all strategy before the start of play cannot be maintained as a descriptive theory simply for the reason that there are far too many possible strategies to be considered The players may use *strategic considerations* which guide their play, but in general they do not pick a rule which specifies what action they should take in any conceivable situation which might arise, much less survey all possible such rules before making their decision

In order to represent formally the essential features of decisions in games, it is necessary to extend the basic formalism of the general decision situation Instead of a single set of acts, there will be sets of acts for each of the "players" in the game If there are n players in the game, the sets of acts or strategies will be denoted $A_1, A_2, \ldots, A_n$ As before, there will be a set S of states of nature representing all the factors determining the outcome of the game which are not controlled by the players Finally, after each player has chosen a strategy there will be an outcome for each player for any state of nature Therefore, instead of there being a single outcome function, there will be n outcome functions $M_1, \ldots, M_n$, where M_i is the function which determines what the outcome of the game will be to the i^{th} player Formally, for each $i = 1, \ldots, n$, M_i is a function defined over the *cartesian product* of the sets $A_1, \ldots, A_n$ and S, such that if $a_1, \ldots, a_n$ are the strategies chosen by players $1, \ldots, n$, and s is the state of nature, then $M (a_1, \ldots,$

a_n, s) is the outcome to player i One may also define the corresponding payoff functions $U_1, \ldots, U_n$, where $U_i(a_1, \ldots, a_n, s)$ is the utility to player i of the outcome $M(a_1, \ldots, a_n s)$ We shall see below, however, that the introduction of the utilities of two or more people raises some special problems in the conceptual foundations of utility theory not previously encountered

Because the outcome of a game is determined by more than two factors and more than one outcome is included, it is no longer possible to represent decisions in games in the elegant matrix notation developed above There is, however, one exception to the above statement This occurs in what are called 'two person zero sum games In these games the outcome is presumed to depend only on the actions of the two players, plus *random factors* of which the probabilities are known In case the states of nature are all random with known probabilities, then it is possible to eliminate the set S and assume that the actions of players 1 and 2 entirely determine the outcome, which becomes a probability mixture To see why this is so, suppose that for each state s of nature there is a definite probability p_s that it will occur Now, if players 1 and 2 select strategies a_1 and a_2 player 1 will get outcome $M_1(a_1,a_2 s)$ with probability p_s and player 2 will get $M_2(a_1 a_2,s)$ with the same probability Hence, the outcome to player i if actions a_1 and a_2 are taken will be a probability mixture in which he gets $M_i(a_1 a_2 s)$ with probability p_s The utility $U_i(a_1 a_2)$ from these two strategies will simply be the expected value

$$U_i\ (a_1,a_2) = \sum_{s \epsilon S} p_s U_i(a_1,a_2\ s)$$

The elimination of the set S in two person games means that the outcomes depend on only two factors, namely the strategies chosen by the two players, and a matrix representation is suggested, in which the rows correspond to the strategies of player 1 and the columns to the strategies of player 2 Nevertheless, there are still two outcomes, and two utility payoffs, one for each player, and, therefore, it is not yet possible to give a matrix representation, if only one outcome or payoff is to be entered corresponding to each choice of strategies by the players In the so called 'zero-sum" case, however, it is assumed that whatever player 1 wins player 2 loses, and vice versa, and, therefore, the outcome for player 2 is, in a sense, the 'negative of that for player 1, and the utility payoff to player 2 is the negative of that to player 1 In this zero sum case it becomes possible to eliminate the payoff to player 2 from the matrix, and enter only the payoffs to player 1 in the

spaces of the matrix, since those to player 2 can be immediately computed as the negatives of those to player 1 It is customary to represent the decision situation in a two-person zero sum game by a payoff matrix in which the payoffs are the utilities of the outcome to the first player

6 5 2 The Interpersonal Comparability and Transferability of Utilities.

In asserting that in the two-person zero sum game the utility payoff to player 2 is the negative of that to player 1, we have slurred over a conceptual problem of fundamental importance in the measurement of utility This is the problem of giving operational significance to a comparison of the utilities of two different people This problem did not arise until this point because we were previously concerned with the behavior of single individuals in decision situations, and not with situations involving the decisions of two or more people The theory of zero-sum games on the other hand assumes that the gain in utility for one player in a game is exactly equal to the other player's loss in utility If this theory is to be applicable in either a descriptive or a normative sense, some operational significance must be attached to the concept of "equal utility gains or losses" by two different people, or else it must be shown that the theory of two-person zero-sum games does not rest essentially on this concept, and can be reformulated without it In this section, we can only make a few unsystematic remarks about this problem which has, as far as the author knows, never been thoroughly explored

It is to be noticed in connection with the measurement of utility that utility is what is called a "linear scale," which means that if one measurement is defined, it is possible to change it by making an arbitrary change of the unit of measurement, and of the zero-point from which measurement starts But, if one is to give some meaning to the assertion that one person's utility gain is equal to another person's utility loss, it must be assumed that the units of utility measurement are the same for both persons But how is it possible to determine this? One might "solve" the problem by stipulating arbitrarily that the utility gain due to getting one dollar is to be the unit of utility for all people This would lead to the conclusion that getting a dollar is equally valuable to all people — which seems intuitively wrong But, unless some such arbitrary stipulation is made, it is not easy to see how a "rational" definition of "equal utility" between two persons can be given

Fortunately, the theory of two-person zero-sum games does not depend essentially on the direct comparability of utilities It turns out that this

theory only demands that there be *direct opposition* between the interests of the players, in the sense that for any two outcomes x and y, if the first player would prefer x to y, then the second would prefer y to x It is possible to test operationally the direct opposition assumption, and hence the theory is saved both for descriptive and normative application

A problem closely related to that of comparability arises in the theory of n person games (*i e*, games involving more than two people) developed by von Neumann and Morgenstern In this theory it is assumed not only that individual utilities are comparable in the sense described above, but that when groups of individuals get together to form ' coalitions, they can pool their utilities and play as a unit, afterwards dividing up the total utility payoff The assumption that utilities can be pooled or divided up is called the assumption of ' *transferability* The intuitive idea is that utilities can be traded like money The analysis of this assumption is even more difficult than that of comparability, and we can only raise the problem here It is to be noted that, taken literally, the idea that it is possible to transfer utilities is meaningless, or false Clearly what happens in a financial transaction is a transfer of money, not of some sort of subjective quantity One may suppose that under some circumstances the gain in utility to one person in the money transfer is exactly equal to the loss in utility to the other in the transfer, but this brings us back once again to the problem of comparability

6 6 Decision Principles

6 6 1 General Comments

We shall now give a brief discussion of the various types of solutions which have been proposed for the decision problem or for special cases of it The decision problem formally stated simply asks given preference relations P, I, and R among the outcomes O of the decision situation (or, given the utilities of the outcomes), what should be the preferences among the set A of acts? A *decision rule* is one which specifies some connection between the preferences among the outcomes and those among the acts or some relation among the preferences for the acts The strongest kind of decision rule is one which, given the decision situation and the preferences or utilities of the outcomes, determines just which act the set A should be chosen The *minimax* and *minimax regret* rules discussed below are of this type A much weaker type of rule is one which specifies only that certain actions *should not* be chosen, though not which actions *should* be The *dominance* principle described below is of this type Still another type of rule is one which asserts that if certain preferences among acts are held, then certain others ought to

be held, these rules may be thought of as "*consistency principles*" in that they require a kind of consistency in the choices among acts without specifying which act is to be decided upon The *weak ordering* rule, and Savage's *sure-thing* principle of Sections 6 6 2 and 6 6 4, respectively, are examples of consistency principles

The decision rules described in this section may be categorized in still another way The principles discussed in Sections 6 6 2 and 6 6 4 are all necessary conditions for what is called a "*Bayes solution*" of the decision problem A Bayes solution is one in which the person making the decision acts as though he assumes that each state of nature has a definite probability, in which case it becomes possible to compute the expected utility of each act and choose the one with the highest utility This type of solution is discussed in Section 6 6 5, and, therefore, Sections 6 6 2 and 6 6 5 and the principles described there form a unit The decision principles discussed in Section 6 6 6 on the other hand are inconsistent with a Bayes solution of the decision problem They are alike in two respects (1) they are decision principles of the strong kind, which the Bayes principle is not, and (2) none of them has the same immediate rational justification that the principles of Sections 6 6 2 and 6 6 4 have. It is to be noted also that, with the exception of the minimax rule, the principles of Section 6 6 6 are of primary importance in statistical decision situations Finally, in Section 6 6 7, we give a very brief discussion of decision rules in two-person games In the situation in which two persons' interests are involved, one would expect that radically new factors enter the decision problem This expectation is correct, and in the two-person game situation, we encounter for the first time the concept of *equilibrium* which is of the utmost importance in the solution of games In the final part of this section, we shall briefly re examine the decision principles previously discussed in order to compare them with the assumptions of Bernoullian Utility theory (Sections 3 and 4) with the hope that this comparison will provide a more critical insight into the assumptions of Bernoullian Utility theory, and their limitations

6 6 2 Weak Ordering of Preferences.

It was assumed that the preference relations P, I and R represented ambiguously preferences for the outcomes O and the acts A So far nothing has been specified about the properties of the preference relations for these two sets, or about the connection between them If, however, it is assumed that the outcomes can be replaced by their utilities in a payoff matrix, and these utilities satisfy the ordinal and expected utility assumptions (Section

3), then certain properties of the preferences for the outcomes follow immediately Throughout most of this discussion it will be assumed that preferences for the outcomes have the properties following from the hypothesis that there exists a Bernoullian Utility function for them, and seek principles determining the properties of the preferences for the acts One consequence of the assumption that there exists a Bernoullian Utility function for the outcomes is that the preference or indifference relation R is a *weak ordering* of the set O (*) This means that the individual has a preference between any two actions and that these preferences are transitive

A second assumption, and one which is entirely independent of the weak ordering assumption for outcomes, is that R is a weak ordering also of the acts This is a consistency assumption in that it does not specify which of two acts should be preferred to the other, but only that if the individual has certain preferences, then he should *also have* certain others The two weak ordering assumptions are stated formally below

Weak Ordering Principle R is a weak ordering of O and of A i e, for all x and z in A or for all x, y, and z in O,

(1) either xRy or yRx,
(2) if xRy and yRz, then xRz

6 6.3 Dominance.

The weak ordering principle specified something about the form of the relation R in the sets A and O separately, but nothing about the dependence of preferences among acts on preferences among outcomes The dominance principle to be described below does specify one such connection, albeit one which does not determine many preferences among actions The intuitive idea of the dominance principle is that if there are two acts available, one of which has a better outcome than the other, no matter what the actual state of nature then the first act is to be preferred to the second An example of a dominance situation is one in which a person is trying to decide which of two medicines A and B to take to cure an illness the symptoms of which are consistent with its being either one of two diseases x and y If it should so happen that medicine A was more effective in curing both x and y than medicine B, then the alternative of taking medicine A would *dominate* the alternative of taking medicine B in this case and the individual would have no hesitation in choosing medicine A

(*) See Consequence 1, p 185

Dominance Principle. If a and b are members of A such that for all members s of S, $(M(a,s))R(M(b,s))$, then aRb

It is very seldom that the dominance principle by itself guides the individual to a decision in a decision situation Practically, its function is negative, in that it eliminates from consideration any actions which are dominated by other actions This principle, however, gives no guidance in *conflict situations*, in which the consequences of one act are better in some states of nature, but are worse in others than those of another action Most important decisions are of the conflict variety and, hence, there is a need to develop principles of decision in these situations

6 6.4 Savage's Sure-thing Principle.

The dominance principle is often referred to as a "sure-thing principle," because it says (in effect) that if one act is better than another in its consequences, *no matter what*, then it should be chosen over the other A particular instance of this kind of situation is the "sure thing" bet in which the bettor is sure to come out ahead no matter what happens L J Savage (*Foundations of Statistics* [1954]) has stated a more elaborate decision principle, which he also calls "the sure-thing principle," which is related to the dominance principle, though it is not equivalent to it The intuitive idea of Savage's principle is as follows Suppose that there are two actions a and b which have identical outcomes in certain states of nature, and such that a is preferred to b For example, a might be to bet \$10 on football team X to win a certain game against Y, and b might be to bet \$10 on Y to win in this game, with the condition that the bet is cancelled in case of a tie There are three "states of nature" in this situation X wins, or Y wins, or they tie Actions a and b have the same outcome in the case of the third state of nature in which there is a tie, since in either case the person neither wins nor loses Now, suppose that acts a and b are changed to a' and b', respectively, only in the outcomes in which a and b are the same, and both are changed in the same way Then it is stipulated that if a is preferred to b, then a' should be preferred to b' In the example of the man betting on the football game, acts a and b in which the bet is called off in the event of a tie might be changed to a and b', exactly like a and b except that in the case of a tie the bettors decide to flip a coin to determine who gets the \$10 Savage's sure-thing principle then stipulates that if a man prefers act a (betting \$10 on team X) to act b (betting \$10 on team Y), in which the bet is called off if a tie occurs, then he should prefer a to b since a and b' are exactly like a and b except

that in the one case the bet is called off if a tie occurs and in the other they flip a coin for the \$10 The situation involved here is represented in the following outcome matrix

	X wins	Y wins	Tie
a	win \$10	lose \$10	get \$0
b	lose \$10	win \$10	get \$0
a'	win \$10	lose \$10	flip coin
b'	lose \$10	win \$10	flip coin

The intuitive idea of Savage's sure thing principle is simply that the decision between two acts a and b should depend on the outcomes which they have which are different, so that if a and b are changed only in the outcomes which they have which are the same, the order of preference between them should not change This principle is stated formally below Let a, b, a', and b' be members of A, and T be a subset of S such that (1) for all s in T, $M(a,s) = M(b,s)$ and $M(a',s) = M(b',s)$, and (2) for all s in $S - T$, $M(a,s) = M(a',s)$ and $M(b,s) = M(b',s)$ Then, if aRb, then $a'Rb'$

In the statement of Savage's sure thing principle above, *T is the set of nature in which a and b have the same outcomes — in the example, T includes just the one state of nature in which a tie occurs* $S - T$ is the set of all the other states of nature, *including in the example* the event of X winning and the event of *Y winning*

Savage's sure *thing principle* is not as intuitively sure a principle as the dominance principle, though in Savage's theory the dominance principle *is derived from* his sure thing principle plus the assumption that R is a weak *ordering* It will be seen later that there are decision theories in which Savage's sure thing principle does not hold

6 6 5 Bayes Solutions

Each of the three decision principles discussed in Sections 6 6 2 and 6 6 4 may be thought of as consequences of the *Bayes decision principle* to be discussed next The intuitive idea of the Bayes principle is that *in forming* his preferences among the acts the individual must act *as though* he forms estimates of the probabilities of the states in S and that his preferences are in accordance with the expected utilities derived from these probabilities If

each state s in S is assigned a probability p_s. Then the outcome of any action a would simply be a probability mixture of the outcomes $M(a,s)$, in which $M(a,s)$ occurs with probability p_s. The expected utility of this mixture is then simply the sum

$$\sum_{s \in S} p_s U(a,s)$$

If the utilities $U(a,s)$, are already defined, then the Bayes principle specifies that the individual must assign probabilities p_s and make his choices among acts accordingly. If only the individual's preferences among outcomes are defined, then the Bayes principle requires that he assign both the utilities and the probabilities and make his choices among acts accordingly. Each of these two principles is formalized below.

First Bayes Principle. There exists a probability distribution p over the set S such that, for all a and b in A, aRb if and only if

$$\sum_{s \in S} p_s U(a,s) \geq \sum_{s \in S} p_s U(b,s)$$

Second Bayes Principle. There exists a probability distribution p over S, and a real valued function u defined over O such that, for all a and b in A and s and t in S,

(1) $(M(a,s))R(M(b,t))$ if and only if $u(M(a,s)) \geq u(M(b,s))$,

(2) aRb if and only if

$$\sum_{s \in S} p_s u(M(a,s)) \geq \sum_{s \in S} p_s u(M(b,s))$$

It is clear that the second Bayes principle is very much like the theory of subjective probability discussed in Section 5.3. It is not immediately evident from the statements of the Bayes principles just what behavioral implications they have and, therefore, the problem of finding the empirical consequences arises here, as for Bernoullian Utility theory. Each of the decision principles previously described are consequences, but these do not exhaust the implications of the Bayes principles for behavior. Savage in his book *Foundations of Statistics* has given a set of behavioral postulates which are equivalent to the second Bayes principle above.

The decision theory represented by the Bayes principle, which includes all of the previous principles, is weak in the sense that, though it specifies that the individual assign probabilities p_s to the states of nature, it does not stipulate how these probabilities are to be assigned. Obviously a change in

the assignment of probabilities makes a great deal of difference in the preferences made in accordance with those assignments This weakness in the Bayes theory indicates a fundamental limitation on all of the principles discussed so far, since all are included in the Bayes principle In the following section we shall consider some principles which do not have this weakness, and, in fact, given any decision situation with a payoff matrix, they will determine the action to be chosen uniquely

6 6 6 Minimax and Minimax-regret.

The minimax and minimax-regret principles are of the kind that do not give a complete ordering to the alternative acts in the set A, but only specify, given the entire set, which should be chosen As we shall see, however, it is easy to modify these rules in a natural way to give a complete ordering Intuitively the minimax rule says that the individual should always choose that action in which the possible loss is minimized In other words, this principle says to choose as though, no matter what act is chosen, the worst outcome will happen Then the person will choose that act with the best "worst outcome " If there are two or more such acts, then the individual may choose between them on other grounds

The principle can be more easily stated formally, if it is extended to give a complete ordering of the acts One act a is to be preferred to a second act b if the worst that can happen under a is better than the worst that can happen under b

Minimax Principle. For all a and b in A, aRb if and only if there is some s in S such that for all t in S, $(M(a,t))R(M(b,s))$

Essentially, what the above principle says is that a is preferred to b if there is some state s such that the outcome $M(b,s)$ of action b in state s is worse than all outcomes $M(a,t)$ in any state t

There are many reasons why the minimax principle does not seem to be a reasonable one Intuitively, what it directs anyone to do is take the course dictated by extreme pessimism, and there seems little reason to regard this as "rational " Another way to test the rationality of the minimax principle is by determining how it accords with the three principles set forth in Sections 6 6 2 and 6 6 4, since each of these can be justified as a postulate of rationality It turns out that the minimax rule is consistent with both the requirement that R be a weak ordering of the acts (at least in case S is a finite set), and with the dominance principle It is not, however, consistent with Sa-

vage's sure thing principle, as the following example shows Suppose there is a decision situation with two states of nature s and t, and four acts a, b, a', and b', with the following outcome matrix

	s	t
a	\$10	—\$10
b	—\$10	—\$10
a'	\$10	\$10
b'	—\$10	\$10

If, as may be assumed, getting \$10 is preferred to losing \$10 (denoted "—\$10" in the outcome matrix), then, according to the minimax principle aIb, since the worst outcome in both cases is —\$10 On the other hand, the preference between a' and b' should be $a'Pb'$, since the worst outcome under a' is getting \$10, but the worst possible outcome of b' is —\$10 In this situation, however, Savage's sure-thing principle asserts that if aIb, then $a'Ib'$, contradicting the minimax rule Other interesting discrepancies occur between the minimax and the Bayes principles, which there is not space to consider here (*) All of these "contradictions" have the effect of casting doubt on the minimax rule as a principle of rationality Nevertheless, as we shall show in the following section, the minimax rule can be given a strong justification when applied to game situations

An interesting variation of the minimax principle is the "*minimax-regret*" principle, originally proposed by L J Savage [1954] In this case the individual chooses that act which has the least possible "regret " If an individual chooses act a and the state of nature is s, then the "regret" in this situation is defined as the difference between what he actually gets (namely $M(a,s)$) and what he might have gotten if, with the same state of nature, he had chosen the best possible act

It is clear that it is not possible to formulate the minimax regret principle directly in terms of the outcome matrix and preferences among the outcomes, since regret is defined as the *difference* in value of two outcomes, which depends on a utility measure The *regret function* R for the outcomes and states must, therefore, be defined in terms of the payoff function U for all a in A and s in S,

$$R(a,s) = \max_{b \epsilon A} (U(b,s)) - U(a,s)$$

(*) More extended discussions are given in Radner and Marschak [1954] and Chernoff [1954]

In the formula above, $\max_{b \epsilon A} (U(b,s))$ is simply the maximum utility payoff to the individual under state of nature s. Now, the minimax regret principle says that the individual should choose that action for which the possible regret is a minimum. This can be formulated as a rule which specifies the preference between any two alternatives, by stipulating that for any two alternative a and b, a should be preferred to b if the maximum possible regret under a is less than that under b.

With the regret function R defined as above, the minimax regret principle is formulated as follows:

Minimax-regret Principle For all a and b in A, aRb if and only if

$$\max_{s \epsilon S} (R(a,s)) \leq \max_{s \epsilon S} (R(b,s))$$

The minimax-regret principle violates the same decision principles as the regular minimax rule, *i.e.*, it violates the Savage sure thing principle. However, it also violates the dominance principle, as the following example shows. Suppose that there are two actions a and b, and two states of nature, s and t. The payoff matrix in the example is

	s	t
a	4	2
b	1	0

In this case the maximum possible regret under action a is 2, and the maximum under b is 1, therefore, the minimax regret principle specifies that b should be chosen over a although action a actually dominates b.

6.6.7 Solutions to Two-person Games.

As would be expected, rational decision principles in games differ considerably from those in other decision situations, in which the only factors which influence the outcome of the decision are the states of inanimate nature, which is not supposed to have "interests" of its own. The chief difference between the situations in which the only factors besides the action of the individual influencing the outcome are the "states of nature" and those in which other self-interested individuals are involved is that these other individuals may attempt to anticipate the given person's action, in order to take the maximum advantage of it. In such a case, the person in making

up his mind as to what strategy to choose can *no* longer presume that the other factors influencing the outcome of the decision are "independent" of his own decision. If the state of nature is thought of as the strategy chosen by nature in deciding the outcome of the ordinary decision situation, the player can at least assume that nature does not try to outguess *him* in choosing it.

In the case of the general game, there are no decision principles which have so far been advanced which everyone agrees upon as being "rational." Even the dominance principle which seems as rational as any leads to rather paradoxical situations in some cases, of which the following is an example. Suppose that two players play a game in which each has two strategies, a_1 and b_1 for player 1, and a_2 and b_2 for player 2, and once the players pick their strategies, the outcome is completely determined, since it does not depend on a state of nature. The payoffs are determined according to the following rule: if a_1 and a_2 are picked, then neither player gets anything; if a_1 and b_2 are picked, then player 1 gets $11, and player 2 loses $1; if b_1 and a_2 are picked, then player 1 loses $1, and player 2 gets $11; finally, if b_1 and b_2 are picked, then both players get $10. This situation can be represented in the following matrix type formulation, in which the outcomes to both players are entered in the spaces of the matrix, those on the upper left part of the rectangle being to player 1, and those in the lower right being to player 2.

Player 1's acts	Player 2's acts: a_2	Player 2's acts: b_2
a_1	$0 / $0	$11 / —$1
b_1	—$1 / $11	$10 / $10

In this case the dominance principle dictates that player 1 should use strategy a_1, and player 2 should use strategy a_2, in which case both players get nothing, though if they chose b_1 and b_2, respectively, both would get $10.

There is one situation in game theory in which there is a generally agreed upon decision principle. This is in the two-person zero-sum situation mentioned briefly in Section 6.5.2. The rule is that each player should use a minimax strategy, when the two players' strategies are what are called "*equilibrium*" strategies. Two strategies a_1 and a_2 by players 1 and 2, respectively, are

such as the above one, in which there is no equilibrium pair in the set of pure strategies, the theory of two person games asserts that the players should *prefer* mixed strategies to pure ones In this example, the solution is that each player should use a random device in choosing his strategy which picks each with 50 % probability An intuitive justification for the use of mixed strategies in these zero sum games is that it is to the player's interest to insure that the opponent will not be able to guess which strategy he will use, and that the use of a random device for picking this strategy will make this impossible A second justification is that, if a series of such games as in the above example are played, then it would be very foolish for one player consistently to choose the same act, since if the other player found this out, he could always make sure of winning Hence, the player should vary the strategy he chooses according to a random pattern

6 6 8 Comments on the Assumptions of Bernoullian Utility Theory. We shall end this chapter with a comparison between the decision principles of Bernoullian Utility theory, and some of those discussed previously in this chapter The first thing to be noted in this connection is that there is no conflict between Bernoullian Utility and the Bayes principle and, therefore, no conflict between Bernoullian Utility and any of the principles discussed in Sections 6 2 to 6 5 The reason for this is that Bernoullian Utility can be thought of as the special case of the Bayes theory in which the individual assigns to the events their objective probabilities Hence decisions made in accordance with Bernoullian Utility satisfy the decision principles implicit in the Bayes principle, *plus* the principle that the individual must assign the objective probability to the event By converse reasoning, anything which violates the Bayes principle (or any of the principles of Sections 6 2 to 6 4) violates the assumptions of Bernoullian Utility theory Hence the minimax and minimax-regret principles both violate the assumptions of Bernoullian Utility theory The important thing is to determine just which of the empirical consequences of the expected utility hypothesis do not hold if the individual uses either the minimax or minimax-regret principles

There are two important consequences of the expected utility hypothesis which do not hold for persons using either the minimax or minimax-regret principles These are Consequences 3 2 and 3 4 Consequence 3 2 asserts that if $p > 0$ and xPy, then for any z, $\langle px, (1-p)z \rangle \; P \; \langle py, (1-p)z \rangle$ Suppose that x = winning $10, y = breaking even, and z = losing $10, and p is the probability that a fair coin will fall heads, *i e*, $p = 5$ Then, clearly xPy and, according to Consequence 3 3, it should be that $\langle 5x, 5z \rangle$

$P < 5y,\ 5z>$ But if this situation is set up in matrix form, the minimax principle requires that $< 5x,\ 5z>\ I < 5y,\ 5z>$ In the matrix form the two acts in question are $a = < 5x,\ 5z>$ which is simply the action whose outcome is to get $10 if the coin falls heads and lose $10 if it falls tails, and $b = < 5y,\ 5z>$, which is to get nothing if the coin falls heads and lose $10 if it falls tails The two states of nature are simply the coin's falling heads (H) and its falling tails ($\bar{H}$)

	H	$\bar{H}$
a	$10	—$10
b	$ 0	—$10

Clearly the minimax principle specifies in this case that aIb, since both have the same "worst outcome"

Corollary 3 4, which is a consequence of Consequence 3 2 and the mixture space axioms, is violated in the same way It says that if xPy and $p > q$, then the mixture $<x,p(1-p)y>$ should be preferred to the mixture $<q\,x\,(1-q)y>$ According to the minimax principle, if q is less than 1, then $<px,(1-p)y>$ should be indifferent to $<qx,(1-q)y>$ since both have the same worst outcomes, namely, y

If the minimax principle can ever be taken as a rational decision principle, then in those instances in which the minimax principle is rational, Consequences 3 2 and 3 4 must not be rational One case of primary importance in which the minimax principle is rational is in zero sum two-person games This then indicates a limitation on Bernoullian Utility theory it cannot be applied to acts in which the outcome of the action is itself a strategy in a game, because the rational decision principle in these situations is minimax, which is inconsistent with Bernoullian Utility This is a very important restriction on the applicability of Bernoullian Utility to ordinary decision situations because the result of any decision will as a rule, not only include the acquisition of certain goods, but also will have the function of putting the individual in a certain strategic position Thus, a man may risk a certain sum in a gamble with the hope of increasing his wealth, but also with the hope that with this newly gained wealth he may take bigger gambles

7 EMPIRICAL APPLICATIONS OF BERNOULLIAN UTILITY THEORY

7.1 Preliminary Remarks.

The empirical application of Bernoullian Utility theory has been largely transplanted from the domain of economics to that of psychology in the ten years since it was first revived by von Neumann and Morgenstern This trend is reflected in the fact that while the first papers suggesting empirical applications of Bernoullian Utility to come out following the appearance of the *Theory of Games and Economic Behavior* applied this theory to the explanation of certain economic phenomena, most of the recent papers on applications have been concerned with experimental tests of the assumptions in explaining individual choice behavior, and have been written by psychologists Probably the reason for this shift is the fact that it is possible to get a grip on the central concepts of the theory — preference and utility — much more directly in the controlled situations in psychological experiments than is possible in the relatively uncontrolled economic explanation of mass phenomena Utility theory in economics has traditionally served as an intuitively satisfying theory of individual choice behavior furnishing an underpinning to the explanation of mass phenomena But the theory is removed at some distance from any direct test, and from one point of view is quite unnecessary to the mass theories built on it Modern trends in economics have tended to discard individualistic explanations of mass phenomena The trend away from individualistic utility hypotheses as explanations in economics seemed to be momentarily reversed with the advent of the von Neumann Morgenstern theory, since this theory provided a long sought "cardinal" measure of utility However, the basic trend has continued, and there have been relatively few serious attempts to apply Bernoullian Utility in explaining observed mass phenomena in the past five years

In this section we shall discuss three applications of Bernoullian Utility theory, one from the domain of economics and two from psychology The economic application is Friedman and Savage's and Markowitz's theory as to why people gamble, and why they buy insurance — both types of behavior involving a statistical expectation of monetary loss The theory is advanced as an explanation of certain types of behavior which are almost universal but which are not in accord with the assumption that people try to maximize monetary gain Although the Friedman Savage and Markowitz theories do not incorporate any experimental tests, they propose certain hypotheses about the utility of various amounts of money which suggest

such tests The second application is an experiment by Mosteller and Nogee to test the Bernoullian Utility hypotheses as applied to situations in which the alternatives are amounts of money or "gambles" whose outcomes are amounts of money The final application is likewise an experimental test of Bernoullian Utility as applied to money alternatives and mixtures, but this time incorporating a theory of subjective probability, and certain refinements of technique which are not included in the Mosteller Nogee experiment

Besides the two experiments which we discuss there have been a great variety of experimental tests of various aspects of Bernoullian Utility and its variations It is impossible to discuss all these here, and the two experiments discussed must be taken as representative Of more significance is the fact that a detailed discussion of these experiments illustrates more clearly than any purely theoretical exposition the great difficulty of designing a significant experiment to test the apparently simple Bernoullian Utility hypotheses

7.2 Hypotheses Explaining Gambling and Insurance Buying.

Two papers by Friedman and Savage [1948 and 1952] and one by Markowitz [1955] propose hypotheses to explain why people gamble and buy insurance The central fact of somewhat paradoxical nature in both gambling and the buying of insurance is that the expected value of the money return in both instances is negative (at least the expected return is negative in gambling where there is a house "cut," as in roulette) The classical theory of gambling assumes that persons should always choose the alternative which has the highest expected monetary value Clearly the individual who gambles or buys insurance could choose a course with a higher expected money return by simply not gambling or not buying the insurance In modern theories of decision under risk, the person is supposed to choose that alternative which has the highest expected *utility* outcome, hence there is no contradiction with current utility theory in the fact that a person may not act to maximize expected money, and one may attempt, as Friedman and Savage and Markowitz do, to explain gambling and insurance buying by assuming that the utility of money is not proportional to the amount

Before proceeding to their theories, let us note that the fact that people do not play to maximize money, and that in some instances it seems utterly irrational to play this way, was noted in the 18th century in connection with the *St Petersburg paradox* which led Daniel Bernoulli to propose the first

"Bernoullian" Utility scale The St Petersburg paradox concerns a game which is played in the following way The "house" allows the player to toss a fair coin as many times as necessary until it falls heads (say n times), then the house pays the player 2^n dollars The question is, how much should the house charge the player to play this game? If the house is interested in making sure that its own expected money return is positive, then it should charge an amount slightly in excess of the expected value of the money to the player from playing the game Conversely, if the player is interested in maximizing the expected value of money return, he should be willing to pay any amount less than the expected value of the money return from the game for the privilege of playing it However, it is easy to show that the expected value of money from this game is infinite, hence the player should be willing to pay any amount of money for the privilege of playing it But to most people, even twenty dollars would be too high a price to pay, the chance of even getting back the amount bet would be just one in two thousand Several ingenious solutions were given to the paradox, most of them preserving the principle that the player should act to maximize his money expectation

Daniel Bernoulli, however, made the (for then) radical suggestion that people do not attempt to maximize expected monetary gain, but rather expected utility gain, and that the utility of money is not proportional to its amount Bernoulli's "resolution" of the paradox consisted in fact in suggesting that the utility of money is proportional to the logarithm of the amount This is, so far as we know, the first instance of an expected utility hypothesis being used to explain behavior in risk situations, and is why we have chosen to baptize the modern theory with Bernoulli's name

Even with Bernoulli's hypothesis, it is possible to modify the game in such a way that its expected value (or utility, in modern terms) is infinite *i e*, if the house pays not 2^n but 2^{2^n} dollars to the player if he tosses the coin n times before it falls heads, the value is then proportional to 2^n, and the expected value is infinite In general, if the utility of money can be arbitrarily large, then it is possible to define a variant of the St Petersburg game for which the expected value is infinite, and for which, therefore, the player should be willing to pay any amount to play Since it seems unreasonable to be willing to pay an arbitrarily large amount to play any game, it can be argued that if the utility of money can be defined consistently at all, then it must be bounded above *i e*, if u is the utility function, and $u(x)$ is the utility of x dollars, there must be some number, say k, such that for all $x, u(x) \leq k$

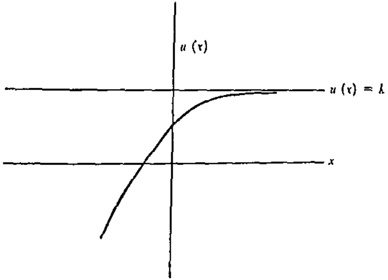


Figure 1

If the function u is plotted graphically with x (the amount of money) on the horizontal axis, and $u(x)$ on the vertical axis, the above argument implies that there is a line above which *the curve does not go* (see Fig 1)

The theories of Friedman and Savage and Markowitz can be interpreted as giving other arguments like the above as to why the utility *vs* money curve should have certain properties

The first *phenomenon* which Friedman and Savage attempt to explain is gambling They take as a typical case gambles in which there is a fairly small probability of winning a large amount, and a large probability of losing a small amount Slot machines, roulette, and lotteries are among this type of gamble All these games have the feature that the mathematical expectation of money winnings in playing them is negative, and is in fact measured by the "house percentage" *Nevertheless, it is* the case that people play them, and, *even leaving* aside the factor of excitement of participation (which we ruled out of consideration in our discussion of the axioms of utility), we may seek an explanation in terms of utility

A typical *gamble* of the type referred to above may be analyzed as follows Let L be the amount that the man bets (*i e*, L is the *amount he* will lose if he loses), let W be the amount *(net)* that the man will win if he wins, and let I be the *amount* of money he has at present Finally, suppose that p is the probability he has of winning — hence $1 - p$ is the probability of losing Then the alternative of betting *is the mixture* with probability p of

yielding $I+W$ (the total amount he will have if he wins), and $1-p$ of yielding $I-L$ (the amount he will have left if he loses) If B denotes this mixture, then

$$B = <p(I+W), (1-p)(I-L)> \ (*)$$

According to the expected utility hypothesis the utility of B is

$$u(B) = pu(I+W)+(1-p)u(I-L) \tag{1}$$

The alternative to gambling is not gambling, which has a certainty of yielding the amount I, which the man has at present If the man prefers to gamble, then

$$u(B) > u(I),$$

or

$$pu(I+W)+(1-p)u(I-L) > u(I) \tag{2}$$

On the other hand, it was supposed that the expected money value of the bet is negative To be more precise, the expected amount of the money the man will have after the bet is less than the amount he had before it If b is expected value, then

$$b = p(I+W) + (1-p)(I-L) \tag{3}$$

It is assumed that b is less than I

$$b < I,$$

or

$$p(I+W) + (1-p)(I-L) < I \tag{4}$$

In Figure 2 this situation is represented graphically Here, as in Figure 1, amounts of money are plotted on the horizontal axis and utility values are plotted on the vertical axis The four money values $I-L$, b I and $I+W$ are located in increasing order along the horizontal axis Note particularly that $I-L$, b, and I are fairly close together, and $I+W$ is far out to the right, this corresponds to the assumption that the amount bet, L, is small compared to the value of W of winning The four utility values $u(I-L)$, $u(I)$, $u(B)$,

(*) Here it is essential to keep in mind the fact that combinations $p(I+W)$ and $(1-p)$ $(I-L)$ in the mixtures are not arithmetical products Below, in computing the expected money values the same combinations will occur as products

and $u(I+W)$ are located in increasing order along the vertical axis, corresponding to the assumption that winning is most preferred, followed by gambling (alternative B), followed by not gambling (alternative I), and last by the certainty of losing (alternative $I-L$)

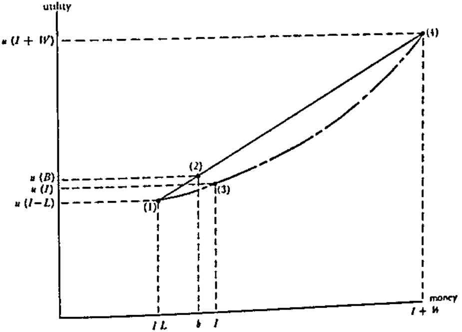


Figure 2

The most significant feature of this graphical representation is the following If the points (1), (2), and (4) are plotted as shown, at the intersections of the vertical lines drawn through the money values on the horizontal lines drawn through the corresponding utility values on the vertical axis, then they must lie in a straight line The reason for this is that the point b on the horizontal axis is the same proportion of the distance between $I-L$ and $I+W$ as the utility value $u(B)$ is between the utilities $u(I-L)$ and $u(I+W)$ (this fact follows from equations (1) and (3)) On the other hand, since $I > b$ and $u(I) < u(B)$, it must be the case that point (3) corresponding to the utility value of I lies below the straight line between points (1) and (4) Hence, if a curve is drawn plotting the utility values of pure money outcomes, $I-L$, I, and $I+W$, it must be concave upwards, as shown by the dashed line in Figure 2

Gambling with negative expected money outcome is now "explained" by assuming that, at least for individuals who gamble with a large prob-

ability of a small loss L and a small probability of a large win W, the curve of utility against money is concave upwards in an interval between I and $I+W$, as shown in Figure 2 This hypothesis says nothing about the form of the utility *vs* money curve outside the region between I and $I+W$, and one may hope to explain other types of behavior involving financial risk by hypotheses about the form of the utility curve outside the interval from I to $I+W$ Friedman and Savage propose the hypothesis that the utility curve is concave downwards (in the opposite direction) in the interval below I (present income) in order to explain the buying of insurance

The insurance buying situation is typified by paying a certain sum P (the premium) to avoid a risk of losing, which for the purpose of this analysis can be represented as a loss in dollars of an amount D The alternative of buying the insurance can be represented as the certainty of having the amount $I-P$ (which is what is left after paying the insurance premium) The alternative of not buying the insurance is a risk, in which there is some probability p that the person will suffer a loss of D, the result of which is that he ends up with $I-D$ dollars, and a probability $1-p$ that no loss will occur, in which case he will end up with what he has at present, namely, I dollars If we again let B denote the risky alternative, then

$$B = \langle p(I-D),\ (1-p)(I)\rangle,$$

and according to the expected utility hypothesis,

$$u(B) = pu(I-D)+(1-p)u(I)$$

If the person elects to buy the insurance, this means that the certainty of getting $I-P$ is preferred to the risk B, hence

$$u(I-P) > pu(I-D)+(1-p)u(I)$$

In the case of buying insurance, the statistical expected value of money from buying insurance is less than the expected value of money in the alternative of not buying the insurance — if it were not, insurance companies would go out of business Since the outcome of buying the insurance is certain, namely, to end up with a total of $I-P$ dollars, the expected value from buying the insurance is simply $I-P$ On the other hand, the alternative of not buying insurance is the risk $B = \langle p(I-D),\ (1-p)(I)\rangle$, and its expected money value is

$$b = p(I - D) + (1 - p)I$$

In this case, b is greater than $I - P$, hence

$$I - P < p(I - D) + (1 - p)I$$

The situation described above is represented graphically in Figure 3

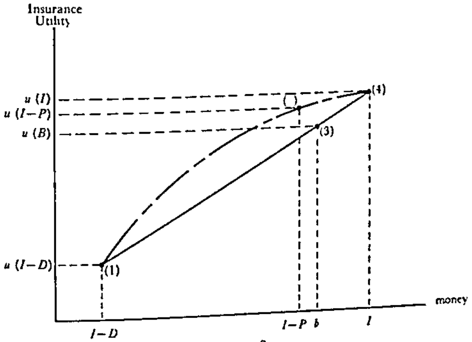


Figure 3

The four money values $I - D$, $I - P$, b, and I are located in increasing order on the horizontal axis, and the four utility values $u(I - D)$, $u(B)$, $u(I - P)$, and $u(I)$ are located in increasing order on the vertical axis. The points (1), (3), and (4) in this case represent the utility values of $I - D$, B, and I, respectively, and, as in Figure 2, they are shown lying in a straight line. Point (2), representing the utility of $I - P$, must lie above this straight line since $I - P$ is less than b, but $u(I - P)$ is greater than $u(B)$. The conclusion is that the utility curve, shown dotted in Figure 3, must be concave upwards in the interval below present income I, down as far as $I - D$.

In Figures 2 and 3, we have shown that utility curves which are concave upwards to the right of I explain gambling, and curves which are concave downwards to the left of I explain insurance buying. We can combine

e into a single curve which explains both However, before constructing final curve, it is well to recall the discussion of the St Petersburg para-, in which it was argued that the utility curve must be bounded above e curve resulting from combining these other curves is shown in Figure 4

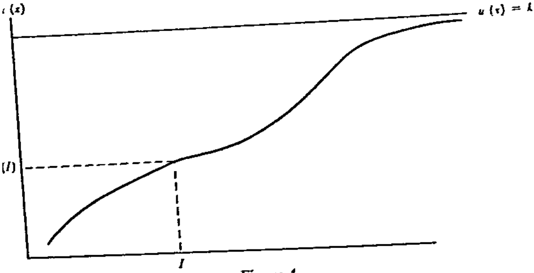


Figure 4

This curve seems to be the simplest type which is consistent with all of the facts discussed so far

It is worthwhile to pause here and see if the curve thus drawn explains any other well known facts other than the ones which it was originally constructed to explain Friedman and Savage consider the factors influencing the distribution of prizes offered in lotteries They note that almost all lotteries offer a graded series of prizes, starting with one or two very large prizes at the top, and working down to quite a few rather small prizes They assume that the lottery operators attempt to construct the schedule of prizes in such a way that their profit from the lottery is a maximum subject to the restriction that the customers regard the tickets as worth the purchase price This phenomenon can be translated into utility terms in which lottery tickets represent risk outcomes with risk utilities which depend on the prizes offered and the probabilities of winning them, and the lottery operator seeks to adjust the prizes and probabilities in such a way that the utility of a ticket is greater than the utility of the purchase price, and at the same time the sum of the amount of the prizes is a minimum (and hence his profit is a maximum) It can be shown that if the utility is convex upwards everywhere to the right of *I*, instead of just in an initial interval as shown in Figure 5, then the lottery ticket operator could make the most money by

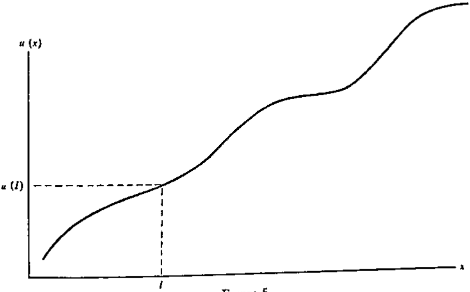


Figure 5

offering just a single very large prize, rather than by offering a number of prizes of varying amounts and varying probabilities Therefore, the fact that lotteries do in fact offer a variety of prizes argues for the fact that the utility curve does not continue to bend upwards indefinitely to the right of I, and must instead start bending the other way again as it moves farther out

Another fact, cited by Markowitz, is that people in general reject "symmetrical" bets, that is, bets in which the amounts that can be lost or won are about the same (this is not supposed to extend to very small bets, in which it can be assumed that the amount of money involved is not important to the bettors) The fact that the curve as drawn in Figure 4 is symmetrical about $u(I)$ provides an explanation of this phenomenon The reader can convince himself of this by representing the amounts to be won and lost at equal distances on either side of I, and connecting the corresponding utility points by a straight line, as was done in Figures 2 and 3 Bets which have a greater than 50 per cent chance of losing will have utilities lying on this line to the left of its midpoint, hence below the x axis, which represents the utility of I Hence these bets will be rejected

Friedman and Savage suggest that the utility curve may in fact be more complex than the one drawn in Figure 4 that it may instead have several "humps," as shown in Figure 5 The curve of Figure 5 still explains all the facts mentioned so far, and there seems no reason to prefer one to the other However, Friedman and Savage suggest that these "steps" may in fact represent discrete levels of aspiration for the individual, corresponding to

definite social classes whose wealth corresponds to the different levels At the top of each hump, there is a certain interval in which a large change in wealth carries little corresponding change of utility Friedman and Savage say that this may be due to the fact that all the incomes in this interval are associated with one economic class, and that a change in wealth as long as one remains in the same class may not be important, whereas a change in wealth which carries a person from one class to another (corresponding to going from one step to another, over one of the steep intervals) may be regarded as much more important

Before passing on to the next topic let us note briefly some possible objections to the theory just presented First, as an explanation of gambling, it leaves out the very important factor of the excitement of participation In the absence of any exact experimental data, it would seem that much of the type of gambling considered in this theory is of the kind in which the amount of money risked is quite small (at least for any one bet), and that the actual value of the money may be of comparable magnitude to the value of the excitement of the gamble One is tempted to surmise that the purchasers of lottery tickets do not do so after sober consideration of the relative values of the money bet and the prizes to be won, but act to a large extent on impulse To argue that the amount of money spent on gambling may amount in total to a sizable portion of the gambler's income and hence that its value is large in comparison to the excitement of gambling is not to the point, since the stipulated interpretation of utility theory requires that it be applied to single decisions

In any event, the principal test which any theory must face is whether or not it succeeds in predicting a large variety of phenomena, and especially phenomena which it was not originally introduced to explain Whether the above theory will meet this test we cannot say, but the criticisms suggest that if it is to be used with any precision, the basic interpretation will have to be more clearly defined In the next sections we discuss two experiments designed to test the theory, in discussing them we shall see some possible ways of giving the basic concepts precise meanings

7.3 The Mosteller-Nogee Experiment.

The experiment of Mosteller and Nogee [1951] can be regarded as an empirical test of the Friedman-Savage theory discussed in Section 7 2 This experiment consisted in running subjects through a series of gambles in which they were permitted either to bet 5¢, or not bet against various amounts of money offered at various odds by the experimenter The game

played was a variety of poker dice in which the experimenter rolled a 'hand of 5 dice and bet a certain sum, after which the subjects (each playing in turn) had the option of betting 5¢ and rolling the dice to try to beat the experimenter's hand, or not betting and passing the dice to the next subject

According to the theory of Friedman and Savage, each subject should possess a 'utility of money curve, and should bet or not according as the expected utility of the bet offered by the experimenter is greater than or less than the utility of no change (*i e*, not betting) For the purposes of this experiment, the zero points of each person's utility scales were fixed at zero cents (*i e*, at their state at the time of the bet), and the unit was chosen so that a loss of 5¢ had a utility of —1 With these two stipulations each person's utility scale is fixed uniquely, and the utilities of every other gain or loss can be measured in terms of the utility of losing 5¢ According to the Friedman Savage theory, once the zero point and unit of measurement have been chosen, to determine the utility of any amount of money, say n cents it is only necessary to find some probability p, such that the subject is indifferent between a bet which offers a probability p of winning n cents and $1-p$ of losing 5¢, and the alternative of not betting If $u(n¢)$ is the utility of n cents, then the utility of a bet $B = <p(n¢), (1-p)(-5¢)>$ which offers a probability p of winning $n¢$ and $1-p$ of losing 5¢ is

$$u(B) = pu(n¢) + (1-p)u(-5¢),$$

and if this is indifferent to not betting with the certain outcome of getting 0, then

$$u(B) = pu(n¢) + (1-p)u(-5¢) = u(0¢)$$

The utility of 0¢ was arbitrarily fixed at 0, and that of losing 5¢ at —1 hence the above equation reduces to

$$u(B) = pu(n¢) + (1-p)(-1) = 0$$

or

$$u(n¢) = \frac{1-p}{p}$$

The general procedure of Mosteller and Nogee was as follows First, for each subject for various amounts of money, say n they determined probabilities p such that the subject was indifferent between not betting and the mixture $B = <pn¢, (1-p)(-5¢)>$ (the behavioral interpretation of indifference will be discussed below) The formula given above was then

used to compute the utilities of these amounts of money for each subject Finally, these computed utility values (which represented points on the subjects' "utility of money" curves) were used to predict choices in new betting situations, and the success of these predictions constituted a test of the theory

It should be noted, of course, that the mere fact that a curve can be plotted using the formula

$$u(n¢) = \frac{1-p}{p}$$

is not evidence tending to confirm the theory Obviously there will be some probability for which the subject is indifferent in this betting situation, and putting that into the above formula, it is possible to calculate $u(n¢)$ in a mechanical way The test of the theory is whether or not the subject chooses alternatives which maximize the expected value of the utilities thus calculated Mosteller and Nogee tried two such tests, applying the information plotted in the original utility curve to try to predict behavior in new situations The first test was to try to predict the behavior of subjects faced with 'doublet' bets, that is, opportunities to make a single bet against two hands at the same time, where it is possible to win either one of two amounts of money, or both, or lose 5¢ This is a different type of situation from that which provided the data on which the curve was based, but if the theory is correct, then the data contained in the plotted curve should predict the subject's behavior in the new situation Hence the new situation furnishes a test for the theory The doublet situation is represented formally as follows Let p_1 and p_2 be the probabilities of beating the first and second hands, respectively (assume that the first hand is higher than the second, hence the probability of beating it is smaller $p_1 < p_2$) and that n_1 and n_2 are the amounts to be won by beating hands 1 and 2, respectively The probability of beating both the higher and lower hands and winning $n_1 + n_2$ cents is p_1, the probability of beating the second hand but not the first hand and winning only n_2 cents is $p_2 - p_1$, and the probability of not beating either and losing 5¢ is $1-p_2$ Hence, the utility of the doublet bet is

$$p_1 u(n_1¢ + n_2¢) + (p_2 - p_1) u(n_2¢) + (1-p_2) u(-5¢)$$

$u(n_1¢ + n_2¢)$, $u(n_2¢)$, and $u(-5¢)$ are all plotted on the utility curve, hence the utility of this bet can be calculated, and if the theory is correct, the subject should take the bet if this utility is greater than 0, be indifferent if the utility is equal to 0, and reject the bet if the utility is less than 0

A second test of the theory is afforded by *paired-comparison* situations The principle idea is that the subject is forced to choose between one of two hands and money bets to bet against Using the utility curve, the utility of each of the two bets offered by the experimenter can be calculated, and if the theory is correct, the subject should choose that bet with the highest utility To describe this situation formally, suppose that the first bet offered by the experimenter is an amount n_1 on a hand which has probability p_1 of being beaten, and the second bet is n_2 on a hand which has probability p_2 of being beaten If the subject bets against either hand, he must wager 5¢, hence the utility of the first bet is

$$p_1u(n_1¢) + (1-p_1)u(-5¢),$$

and the utility of the second is

$$p_2u(n_2¢) + (1-p_2)u(-5¢)$$

All these utilities are plotted on the curve already constructed so it is possible to calculate the utilities of these bets, and see whether the subject does in fact choose that with the highest utility

The actual operational procedure for determining the points on the "original" utility curve was as follows A long series of trials was run during the course of which each subject had many opportunities of betting or not betting against each of the possible hands, and each of a number of offers on those hands made by the experimenter Thus, one of the hands on which the experimenter made bets was four 4's and one 1, and among the many bets offered by the experimenter on that hand was 25¢, and during the course of the series this particular "hand," and the 25¢ bet by the experimenter were offered many times At the end of the series, the proportion of times that a subject accepted a particular offer on a particular hand was calculated for each of the different offers on the hand, and was plotted as shown in Figure 2 Figure 2 shows the amounts offered on the hand on the horizontal axis, and the percentage of times that offer was accepted on the vertical axis It was expected that for a fixed hand and subject, the higher the offer made, the greater the likelihood of acceptance, and that a curve drawn as in Figure 6 would have approximately the "S" shape shown This expectation proved correct in all cases (except for one subject who left the experiment before its completion), although there was considerable variation in the steepness of the slopes of the steps of these curves These curves were plotted for each subject and each hand, and the point at which they crossed the 50 per cent level was taken to be the money offer on the hand for which the subject was

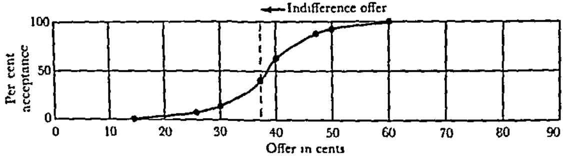


Figure 6

Percentages of times bets of various amounts were accepted by subject X on hand A.

indifferent.(*) For example, in the hypothetical curve drawn in Figure 6, the indifference offer is approximately 37¢. If the probability of beating the hand is p, and the indifference offer is n, then our formula allows us to calculate the utility at n, *i.e.*,

$$u(n) = \frac{1-p}{p}.$$

Thus, each graph like that of Figure 6 for a given subject determines one point on his utility curve, and by plotting these points it is possible to construct a curve like that of Figure 1.(†)

Once the "basic" utility of money curves were plotted, it was possible to test the Friedman-Savage theory by applying the utilities of the basic curves to new situations. It is obvious that the theory cannot be expected to be completely successful in predicting the subject's choices because of the fact that the operational meaning given to "preference" is that the given alternative is chosen more than 50 per cent of the time. But as long as it is possible for a subject to choose an alternative more than 50 per cent of the time, but not all the time, then there must be instances in which he chooses an alternative which he does not prefer, and hence these are instances when the theory predicts he will choose one alternative (the preferred alternative), while he actually picks a different one. The fact that the subject's "*S*" curves, as illustrated in Figure 6, have a non-vertical slope shows that counter-instances exist, in which the subject either chooses to bet, although the experimenter's offer is less than the indifference offer, or chooses not to bet, even though the

(*) Note that indifference is defined here as meaning that each of the alternatives is chosen 50 per cent of the time, similarly, "preference" means that the preferred alternative is chosen more than 50 per cent of the time It is clear that Mosteller and Nogee have tacitly given a stochastic interpretation to the preference relations similar to those discussed in Section 3 5 2.

(†) The possibility that the subjects might not know the true probabilities of beating the various hands was ruled out by providing the subjects with lists giving the objective probabilities

In the experiments to be discussed next some other defects of Mosteller and Nogee's experimental design are commented on

7.4 The Davidson-Siegel-Suppes Experiments.

In the book *Decision Making*, Donald Davidson, Sidney Siegel, and Patrick Suppes [1957] describe a series of experiments which they have conducted to test various aspects of a theory of utility and subjective probability These experiments and the theory which underlies them afford an interesting comparison to the Mosteller-Nogee experiment described in the previous section, in that they subject to explicit test two assumptions implicit in the latter experiment, but not independently tested there These two assumptions are (1) that there is no independent "utility of gambling," and (2) that subjective probability corresponds to objective probability Two more aspects of the Davidson Siegel Suppes experiment are of interest also this experiment gives a different behavioristic interpretation to indifference than was given in the Mosteller-Nogee experiment, and it is based on a finitistic theory which, in the so-called "perfect measurement" case, leads to a unique determination of the utility values of the pure alternatives Because of the complexity of the theory and the experiments it will not be possible to give as detailed discussion of them as was given for the Mosteller-Nogee experiment What we shall do is give a rough informal description of the central ideas of the theory, and of the technique of experimentally testing it

The procedure of the Davidson Siegel Suppes experiments was first to obtain a measure of the utility values of certain amounts of money (based on inferences from observations of choices which involve minimal assumptions about subjective probability), and then, using the utility values thus discovered, to obtain a measure of the subjective probability of certain chance events We shall describe the utility measurement first The theory underlying the measurement of utility is itself divided into two parts a theory of perfect utility measurement, and a theory of approximate measurement The first of these is actually a specialization of the usual theory but applied to a particular finite set of alternatives, the approximate theory incorporates a behavioristic method of interpreting indifference The perfect theory will be discussed first

The Davidson Siegel Suppes experiments and the theory on which they are based deal solely with mixture alternatives of the form $<Ex,\bar{E}y>$,(*) where x and y are amounts of money, and E is some chance event Thus, all

(*) See Sections 3 1 and 5 3 for discussions of mixtures of this form, in which E is the event itself rather than its probability

finite sequence $a_1, \ldots, a_n$ (where $n \geq 5$) amounts of money in the following way. These amounts of money are so chosen that:

$$<a_5E^*, a_1\overline{E}^*> \ I \ <a_2E^*, a_3\overline{E}^*> \tag{4}$$

for $i = 1, \ldots, n-3$,

$$<a_{i+3}E^*, a_i\overline{E}^*> \ I \ <a_{i+2}E^*, a_{i+1}\overline{E}^*>, \tag{5}$$

and for $i = 1, \ldots, n-2$,

$$<a_{i+2}E^*, a_{i+1}\overline{E}^*> \ P \ <a_{i+1}E^*, a_i\overline{E}^*>. \tag{6}$$

Equations (1) — (6) lead to the conclusion that there is some positive number α and some number β such that for $i = 1, \ldots, n$,

$$u(a_i) = \alpha i + \beta. \tag{7}$$

For the alternatives $a_1, \ldots, a_n$ equation (6) gives a measure of utility which is unique once the numbers α and β are fixed, and the choice of these numbers is simply the choice of unit and zero point of utility measurement. The amounts of money $a_1, \ldots, a_n$ may be deliberately selected in order to satisfy condition (4), much as certain weights which are multiples of a fixed unit weight are selected as standards of comparison in determining the weights of other objects in an equal-arm balance. The utility of any other amount of money x can now be approximately determined by finding two adjacent alternatives a_i and a_{i+1} such that $a_{i+1}Px$ and xPa_i. If x lies between a_i and a_{i+1} in preference, then clearly its utility lies between the known utilities of these standard alternatives; hence $u(x_i)$ must lie between $i\alpha + \beta$ and $(i+1)\,\alpha + \beta$.

Even without bringing any new alternatives into the system it is possible to give an independent check with alternatives $a_1, \ldots, a_n$ to determine whether the Bernoullian Utility hypothesis is satisfied by these alternatives and the one chance event E^* with subjective probability 1/2. This check can be carried out by observing all preference relations between alternatives of the form $<a_iE^*, a_j\overline{E}^*>$, where i and j run from 1 to n. The number of possible comparisons is much larger than the original set of comparisons (4) and (5) from which the utility values of the alternatives were derived, yet all possible preferences between these alternatives should be predictable from the knowledge of the utility values of the alternatives if the hypotheses of Bernoullian Utility are satisfied. In fact, it follows from conditions (4), (5), and (6) that for all $h, i, j, k = 1, \ldots, n$,

$$<a_hE^*, a_i\overline{E}^*> \ R \ <a_jE^*, a_k\overline{E}^*> \quad \text{if and only if} \quad h+i \geq j+k. \tag{8}$$

If $n = 6$, then the number of independent preferences which must satisfy condition (8) is 105, and these are predicted from just 10 observations. This procedure illustrates one method of testing a theory involving some unobserved quantities such as the values of the function u: the values of the function are determined by a certain set of observations; then other independent observations are predicted on the basis of these, which, if they are borne out, constitute confirmation of the theory.

So far we have only discussed the theory of perfect measurement. It is obvious that it is critically dependent on the possibility of making indifference judgments, since not only is the unique determination of the utilities of the alternatives $a_1, \ldots, a_n$ via conditions (4)—(6) dependent of these judgments, but the determination of E^* itself as satisfying condition (1) requires indifference judgments. The problem here, as with the Mosteller Nogee experiment, is that there is no direct operational significance to these judgments as there is to strict preferences (determined by observed choices). In order to get around the difficulty raised by these judgments, Davidson, Siegel, and Suppes developed a theory of approximate measurement. Their idea was to arrange the experiment so as to prohibit the subjects from expressing indifference (*i.e.*, they *must* choose one of the alternatives whenever a pair $<xE^*, y\overline{E}^*>$ and $<zE^*, u\overline{E}^*>$ are offered), but to interpret observed choice as *preference or indifference*, rather than as strict preference. The intuitive idea behind this proposal is that if a subject is forced to choose between two alternatives he will choose the preferred one if he has a preference, or pick at random if he has no preference, and therefore if he picks a given alternative it can be inferred that he prefers it or is indifferent between them.

The operational method of determining an event E^* with subjective probability 1/2 can now be explained. It is required that for any two amounts of money x and y, the subject must choose $<(x+1¢)E^*, y\overline{E}^*>$ over $<yE^*, x\overline{E}^*>$, and choose $<yE^*, x\overline{E}^*>$ over $<(x-1¢)E^*, y\overline{E}^*>$. If the subject is observed to make these choices, then it can be inferred that the events E^* and $\overline{E}^*$ are approximately equal in subjective likelihood. For, if E^* were regarded as more likely than $\overline{E}^*$, and we chose $x = 0¢$ and $y = 10¢$, then the subject would probably prefer $<10¢E^*, 0¢\overline{E}^*>$ to $<1¢E^*, 10¢\overline{E}^*>$ since the first would be more likely to give him the 10¢.

Once the event E^* is determined, satisfying the above condition, it is assumed that its subjective probability is equal to that of $\overline{E}^*$, and the measurement of utility proceeds. It is not possible to do more here than indicate some of the salient features of the rather complicated procedure for obtaining an approximate measure of the utilities for the amounts $a_1, \ldots, a_n$ (in

the case discussed in the experiments, $n = 6$) This also depends on getting a series of shifts in preference of the form $\langle (x+1¢)E^*, y\bar{E}^*\rangle$ chosen over $\langle zE^*, w\bar{E}^*\rangle$ but $\langle zE^*, u\bar{E}^*\rangle$ is chosen over $\langle xE^*, y\bar{E}^*\rangle$ These lead to a complicated system of inequalities, which, provided that they are consistent, lead to an approximate value of the utility of each of the amounts $a_1, \quad, a_6$, and these approximations can be improved by bringing in still further observations In this case it is never possible to obtain an exact value of the utility of any alternative so that the best that can be hoped for is a close approximation The situation here is analogous to that which would be the case if one had to determine the weights of certain objects using an equal arm balance which was never level (the balance being level corresponds to indifference) he would only be able to infer certain inequalities among the weights and thus obtain approximate values for them In the actual measurement of utility in the experiment, it was found possible to obtain the utilities of various amounts of money in the range from —50¢ to 50¢ to an accuracy of about one part in ten, *i e*, the range within which the utility value of an alternative must lie (representing what might be thought of as a "maximum error" of measurement) was about 1/10th the utility value of that alternative

Before discussing the results of any of the Davidson-Siegel Suppes experiments, we shall briefly outline the method of obtaining a measure of subjective probability of some chance event E (no longer the event E^* with subjective probability 1/2) using the known utility values of the alternatives $a_1, \quad, a_k$ Let us return again to the case in which the utilities are known exactly, and assume that for the given chance event E and alternatives x, y, z, and w, the indifference judgment $\langle xE, y\bar{E}\rangle \; I \; \langle zE, w\bar{E}\rangle$ is observed If it is assumed that $\Pi(\bar{E}) = 1 - \Pi(E)$, then it follows that

$$\Pi(E) = \frac{u(w) - u(y)}{[u(w) - u(y)] + [u(x) - u(z)]} \tag{9}$$

Hence $\Pi(E)$ is determined by equation (9) if the assumption that the sum of the subjective probabilities of E and its complement are 1 This assumption can be checked in a sense if one is allowed to make the comparison of the alternative $\langle xE, x\bar{E}\rangle$ with the pure alternative x If they are indifferent, then the sum of the subjective probabilities is 1

In case no indifference judgments are permitted, then an approximation method must again be resorted to in order to obtain bounds on the subjective probability of the event E The method used by Davidson, Siegel, and Suppes involved varying the amounts of money in the mixtures by amounts of 1¢

and getting switches in preference indicating the approximate point of indifference, and thus obtaining bounds on the subjective probability.

It is not possible to do more than make a few brief remarks about the actual results of the Davidson-Siegel-Suppes experiments here. Nineteen subjects were tested, and utility curves were constructed for fifteen of them (or rather, a *pair* of utility curves were constructed, representing upper and lower bounds to the approximations). The other four did not satisfy the hypotheses of the theory which were used for the construction of the utility functions. It was observed that for most of these subjects the curve of utility versus money looked like a miniature version of the curve postulated by Friedman and Savage, in the sense that subjects acted as if they would be willing to buy insurance against the possibility of losses, but would gamble with negative actuarial value for the possibility of a gain. The theory was tested by comparing its results with the predictions which could be made assuming the subjects tried to maximize the actuarial values of the alternatives (*i.e.*, the assumption that utility is a linear function of money), and its results were also compared with those of the Mosteller-Nogee experiment. For the fifteen subjects for which utility curves were obtained, all of the predictions which could be made within the limits of accuracy of the utilities measured were borne out, and therefore the Bernoullian Utility theory was, in a sense, "perfect" for these fifteen subjects. One type of comparison was the following: the prediction of the money values x for which subjects would choose the alternative $<-4¢E^*, x\bar{E}^*>$ over the fixed alternative $< 6¢E^*, 11¢\bar{E}^*>$. According to the actuarial theory, subjects should pick the first alternative if x is greater than 21¢, and the second if x is less than 21¢, while no prediction is made if $x = 21¢$. For example, for one of the subjects tested it was predicted that he would prefer the first alternative for x greater than 17¢, and the second for x less than 15¢, contradicting the actuarial theory that he should prefer the second for x between 17¢ and 20¢. In all cases of discrepancy, the subjects obeyed the Bernoullian theory.

In comparison with the results of the Mosteller-Nogee experiment, it was found that in their most general features the utility curves for the subjects agreed, although a direct comparison was difficult since the two curves were derived from different kinds of data by rather different methods.

BIBLIOGRAPHY

Adams, E W and Fagot, R F "A model of riskless choice," Technical Report No 4, [ONR contract Nonr—225 (17)], Stanford University, Aug 1956

Alchian, A "The meaning of utility measurement," *Am Econ Rev*, 43, 26—50 (1953)

Allais, M ' Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école Américaine," *Colloques Internationaux du Centre National de la Recherche Scientifique*, 36 (1952)

— "La psychologie del 'homme rationnel devant le risque La théorie et l'expérience,' *J Soc Stat*, Paris, 93, 47—73 (1953)

— ' L'extension des théories de l'équilibre économique général et du rendement social au cas du risque," *Econometrica*, 21, 269—290 (1953a)

— ' Le comportement de l homme rationnel devant le risque critique des postulats et axiomes de l'école Américaine," *Econometrica*, 21, 503—546 (1953b)

— "On the limitations of the formal method in economics,' *Econometrica*, 22, 58—71 (1954)

Armstrong, W E "Uncertainty and the utility function,' *Econ J*, 58, 1—10 (1948)

Arrow, K J "Alternative approaches to the theory of choice in risk taking situations," *Econometrica*, 19, 404—437 (1951)

Bates, J "A model for the science of decision," *Philosophy of Science*, 21, 326—339 (1954)

Baumol, W J "The Neumann Morgenstern utility index — an ordinalist view," *J Pol Econ*, 59, 61—66 (1951)

Bernardelli, H "A rehabilitation of the classical theory of marginal utility," *Economica*, 19, 254—268 (1952)

Bernoulli, D ' Specimen theoriae novae de mesura sortis," *Comentarii Academiae Scientiarum Imperiales Petropolitanae*, 5, 175—192 (1738)

— ' Exposition of a new theory on the measurement of risk," (English translation of the above by Louise Somer), *Econometrica*, 22, 23—26 (1954)

Blackwell, D and Girshick, M A *Theory of Games and Statistical Decisions*, Wiley and Sons, New York, 1954

Bohnert, H G "The logical structure of the utility concept," *Decision Processes*, R M Thrall, C H Coombs, and R L Davis (eds), Wiley and Sons, New York, 1954

Carter, C F "A revised theory of expectations,' *Econ J*, 63, 811—820 (1953)

Chernoff, H "Rational selection of decision functions,' *Econometrica*, 22, 422—443 (1954)

Coombs, C. H "Psychological scaling without a unit of measurement," *Psych. Rev*, 57, 145—158 (1950)

— ' Mathematical models in psychological scaling," *J Am Stat Assoc*, 46, 480—489 (1951)

— "Theory and methods in social measurement," *Research Methods in the Behavioral Sciences*, L. Festinger and D Katz (eds), Dryden, New York, 1953

— ' Social choice and strength of preference," *Decision Processes*, C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

— and Beardslee, D "Decision making under uncertainty," *Decision Processes*, C H Coombs R L. Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

—, Raiffa, H , and Thrall, R M ' Mathematical models and measurement theory," *Decision Processes*, C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

Corlett, W. J and Newman, P K "A note on revealed preference and the transitivity conditions," *Rev Econ Stud*, 20, 156—158 (1952)

Dalkey, N. "Probability and the measurement of utility,' *Econometrica*, 19, 53—54 (1951)

Davidson, D, Mc Kinsey, J C C, and Suppes, P "Outlines of a formal theory of value," *Philosophy of Science*, 22, 140—160 (1955)

— and Suppes, P. "A finitistic axiomatization of subjective probability and utility,' *Econometrica*, 24, 264—275 (1956)

—, Siegel, S, and Suppes, P *Decision Making*, Stanford Univ Press, Stanford, 1957

Debreu, G 'Representation of a preference ordering by a numerical function,' *Decision Processes*, C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

Edwards, W "Experiments in economic decision making in gambling situations," *Econometrica*, 21, 349—350 (1953a)

— 'Information, repetition, and reinforcement as determiners of two-alternative decisions," *Amer Psychologist*, 8, 345 (1953b)

— "Probability preferences in gambling,' *Am J Psych*, 66, 349—364 (1953c)

— "Probability preferences among bets with differing expected values," *Psych Rev*, 67, 56—67 (1954a)

— 'The theory of decision making," *Psych. Bull*, 51, 380—417 (1954b)

— "The reliability of probability preferences," *Am J Psych*, 67, 68 95 (1954c)

— "Variance preferences in gambling," *Am J Psych*, 67, 441—452 (1954d)

— "An attempt to predict gambling decisions' *Mathematical Models of Human Behavior — Proceedings of a Symposium*, Dunlap and Associates, Stamford, Connecticut, 1955a

— "The prediction of decisions among bets," *J Exp Psych*, 50, 210—214 (1955b)

— "Rewards probability, amount, and information as determiners of sequential two-alternative decisions," *J Exp Psych*, 52, 177—188 (1956)

Egerton, R A D "The holding of assets 'gambler preference' or safety first, *Oxford Econ Papers* (New Series), 8, no 1, (1956)

Ellsberg, D. "Classic and current notions of 'measureable utility',' *Econ J*, 59, 528—556 (1954)

de Finetti, B "La prévision ses lois logiques, ses sources subjectives," *Ann Inst Poincaré*, 7, 1—68 (1937)

— 'Recent suggestions for the reconciliation of theories of probability," *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, J Neyman (ed), Univ of Calif Press, Berkeley, 1951

Flood, M M 'A preference experiment," *RAND Corp Memo*, P 256, Nov 1951

— "A preference experiment," *RAND Corp Memo*, P 258, Dec 1951

— "A preference experiment,' *RAND Corp Memo*, P 263, Jan 1952

— "An experimental multiple choice situation,' *RAND Corp Memo*, P-313, Nov 1952

Fréchet, M "Un problème psychologique sur les probabilités subjectives irrationnelles," *J Psych Norm et Path*, 431—438 (Dec 1954)

— 'Sur l'importance en économétrie de la distinction entre les probabilités rationnelles et irrationnelles,' *Econometrica*, 23, 303—306 (1955)

Freund, R S "Introduction of risk into a programming model,' *Econometrica* 24, 253—263 (1956)

Friedman, M and Savage, L J "The utility analysis of choices involving risk," *J Pol Econ.*, 56, 279—304 (1948)

— "The expected utility hypothesis and the measurement of utility," *J Pol Econ*, 60, 463—475 (1952)

Georgescu Roegen, N "Choice, expectations and measurability," *Quart J Econ*, 68, 503—534 (1954)

Gerlach, M "Interval measurement of subjective magnitudes with subliminal differences," Technical Report No 7 [ONR contract Nonr—225 (17)] Stanford University, April, 1957

Griffith, R M "Odds adjustment by American horse race bettors," *Am J Psych*, 62, 290—294 (1949)

Guilford, J P *Psychometric Methods* (second ed), McGraw Hill, New York, 1954

Hahn, F H "Uncertainty and the cobweb," *Rev Econ Stud*, 23, 65—75 (1955—56)

Harsanyi, J C "Cardinal utility in welfare economics and in the theory of risk taking" *J Pol Econ*, 61, 434—435 (1953)

Hart, A G "Risk uncertainty, and the unprofitability of compounding probabilities," *Studies in Mathematical Economics and Econometrics*, O Lange, F McIntyre, and T O Yntema (eds), Univ of Chicago Press, Chicago, 1942

Hausner, M "Multidimensional utilities," *Decision Processes*, C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

— and Wendel, J G "Ordered vector spaces", *Proc Amer Math Soc*, 3, 977—982 (1952)

Herstein I N and Milnor, J, "An axiomatic approach to measurable utility," *Econometrica*, 21, 291—297 (1953)

Hicks, J R "Robbins on Robertson on utility," *Economica*, 21, 154—157 (1954)

Houthakker, H S "Revealed preference and the utility function" *Economica*, 17, 159—174 (1950)

Hurwicz, L "The theory of economic behavior," *Am Econ Rev*, 35, 909—925 (1945)

Irwin, F W "Stated expectations as functions of probability and desirability of outcomes," *J Personality*, 21, 329—335 (1953)

Jeffreys, Sir Harold *Scientific Inference*, Cambridge University Press Cambridge, 1957

Katona, G "Psychological analysis of business decisions," *Am Econ Rev*, 36, 44—62 (1946)

— "Rational behavior and economic behavior," *Psych. Rev*, 60, 307—318 (1953)

Knight, F H *Risk, Uncertainty, and Profit*, Houghton Mifflin, Boston, 1921

Landahl, H D "A neurobiophysical interpretation of certain aspects of the problem of risks," *Bull Math Biophysics*, 13, 323—335 (1951)

Luce, R D "Semiorders and a theory of utility discrimination," *Econometrica*, 24, 178—191 (1956)

— "A theory of individual choice behavior," Technical Report, Bureau of Applied Social Research Columbia University, May 1957

— "A probabilistic theory of utility," *Econometrica*, 26, 193—224 (1958)

— and Raiffa, H *Games and Decisions*, Wiley and Sons, New York, 1957

McKinsey, J C C *Introduction to the Theory of Games*, McGraw Hill, New York, 1952

Majumder, T "Preference, choice and the theory of games" *Econometrica*, 24, 71—73 (1956)

Malinvaud, E "Note on von Neumann Morgenstern s strong independence axiom," *Econometrica*, 20, 679 (1952)

Manne, A S "The strong independence assumption — gasoline blends and probability mixtures" *Econometrica*, 20, 665—669 (1952)

Markowitz, H "The utility of wealth," *Mathematical Models of Human Behavior, Proceedings of a Symposium*, Dunlap and Associates, Stamford, Connecticut, 1955

Marks, R. W "The effect of probability desirability, and 'privilege' on the stated expectations of children," *J Personality*, 19, 332—351 (1951)

Marschak, J "Neumann s and Morgenstern's new approach to static economics" *J Pol Econ*, 54, 97—115 (1946)

— "Rational behavior, uncertain prospects, and measurable utility," *Econometrica*, 18, 111—141 (1950)

Marschak, J · "Why 'should statisticians and businessmen maximize 'moral expectation'? *Proceedings of the Second Berkeley Symposium on mathematical statistics and probability*, J Neyman (ed), Univ of Calif Press, Berkeley, 1951

— "Toward a preference scale for decision making,' *Readings in Game Theory and Political Behavior*, M Shubik (ed), Doubleday, New York, 1954a

— 'Probability in the social sciences,' *Mathematical Thinking in the Social Sciences*, P Lazarsfeld (ed), Free Press, Glencoe Ill , 1954b

— "Norms and habits of decision making under certainty," *Mathematical Models of Human Behavior—Proceedings of a Symposium*, Dunlap and Associates, Stamford, Connecticut 1955

May, K O · "Intransitivity, utility, and aggregation in preference patterns' *Econometrica*, 22, 1—13 (1954)

Morgan, J N "Can we measure the marginal utility of money? *Econometrica*, 13, 129—152 (1945)

Mosteller, F and Nogee, P 'An experimental measurement of utility,' *J Pol Econ* , 59, 371—404 (1951)

Neisser, H 'The strategy of expecting the worst," *Social Research*, 19, 346—363 (1952)

von Neumann, J and Morgenstern, O *Theory of Games and Economic Behavior* (second edition) Princeton Univ Press, Princeton, 1947

Papandreou, A G , "An experimental test of an axiom in the theory of choice, *Econometrica* 21, 477 (1953) (abstract)

Preston, M G and Baratta, P 'An experimental study of the auction value of an uncertain outcome," *Am J Psych* , 61, 183—193 (1948)

Radner, R and Marschak, J 'Note on some proposed decision criteria,' *Decision Processes*, C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons New York, 1954

Ramsey, F P *The Foundations of Mathematics and Other Logical Essays* Kegan Paul, London and Harcourt, Brace and Co , New York, 1931

Robbins, L 'Robertson on utility and scope," *Economica*, 20, 99—111 (1953)

Robertson, D H *Utility and All That and Other Essays*, George Allen and Unwin London, 1952

— "Utility and all what? *Econ J* , 64, 665—678 (1954)

Roy, A D "On choosing between probability distributions" *Rev Econ Stud* , 22, 194—202 (1954—55)

— "Risk and rank or safety first generalized," *Economica* (new series), 23, 214—228 (1956)

Samuelson, P A *Foundations of Economic Analysis*, Harvard Univ Press, Cambridge, Mass , 1947

— "Probability, utility, and the independence axiom,' *Econometrica*, 20, 670—678 (1952)

Savage, L J The problem of statistical decision,' *J Am Stat Assoc* , 46, 55—67 (1951)

— "Note on Wold s Ordinal preferences or cardinal utility?" *Econometrica*, 20, 663—664 (1952)

— *The Foundations of Statistics*, Wiley and Sons, New York, 1954

Scott, D and Suppes, P "Foundational aspects of theories of measurements,' *J Symbolic Logic*, 23, 113 128 (1958)

Shackle, G L S *Expectations in Economics*, Cambridge Univ Press Cambridge, 1949

— 'The logic of surprise," *Economica*, 20 112—117 (1953)

— *Uncertainty in Economics and Other Essays*, Cambridge Univ Press, Cambridge 1955

— 'Expectation and cardinality," *Econ J* , 61, 211—219 (1956)

Shubik, M 'Information, risk, ignorance, and indeterminacy," *Quart J Econ* , 68, 629—640 (1954).

Simon, H A "A behavioral model of rational choice," *Quart J Econ*, 69, 99—118 (1955)
— *Models of Man,* Wiley and Sons, New York, 1957

Smith, N M, Walters, S S, Brook, F C, and Blackwell, D H "The theory of value and the science of decision," *J Operations Research Soc Amer*, 1, 101—113 (1953)

Sprowls, R C 'Psychological mathematical probability in relationship to lottery gambles," *Am J Psych*, 66, 126—130 (1953)

Stigler, G J 'The development of utility theory,' *J Pol Econ* 58, 307—327, 373—396 (1950)

Strotz, R H 'Cardinal utility,' *Am Econ Rev*, 42 (supplement), 384—405 (1953)

Suppes, P and Atkinson, R C "An analysis of two person games situations in terms of statistical learning theory," *J Exp Psych*, 55, 369 378 (1958)

— and Winet, M "An axiomatization of utility based on the notion of utility difference,' *Management Science,* 1, 259—270 (1955)

Thrall, R M 'Applications of multi-dimensional utility theory," *Decision Processes,* C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

Tintner, G "The theory of choice under subjective risk and uncertainty," *Econometrica,* 9, 298—304 (1941)

Toda, M 'Measurement of intuitive probability by a method of games,' *Jap J Psych*, 22, 39—40 (1951) (abstract)

— "An experimental study on the relationship between two methods of measuring the sequence of values of subjective inference," (English abstract), *Jap J Psych*, 25, 292—293 (1955)

Tornqvist, L 'On the economic theory of lottery gambles," *Skand Actuar Tidskr*, 28, 228—246 (1945)

Tyszynski, H 'Comparisons between equal increments of utility," *Econ J*, 64, 258—263 (1954)

Vail, S V "Alternative calculi of subjective probabilities," *Decision Processes,* C H Coombs, R L Davis, and R M Thrall (eds), Wiley and Sons, New York, 1954

Vickrey, W S "Measuring marginal utility of money by reactions to risk," *Econometrica,* 13, 319—333 (1945)

Wold, H 'Ordinal preferences or cardinal utility?' *Econometrica,* 20, 661—664 (1952)

PART THREE

A Survey of Mathematical Models in Factor Analysis

By HERBERT SOLOMON

STANFORD UNIVERSITY

Contents

1 INTRODUCTION

Factor analysis has a rich and abundant literature In a review of the literature to 1940, with emphasis on the period 1928–1940, Wolfle [1940] presented a bibliography which listed 530 references This author and Rosner [1955] reviewed 164 articles and books published in the three year period July 1952—June 1955 There have been other surveys of the literature which indicate the preparation and publication of a number of papers on factor analysis of the order of magnitude of 1000 since the turn of the century

In the form in which factor analysis is now most frequently encountered, it was motivated by the measurement of mental ability, but it has blossomed in many ways never envisaged by its initial protagonists Within the context of measurement of mental ability, its development began over fifty years ago in a pioneering paper by Spearman [1904] However, factor analysis appeared in embryonic shape some 20 years before that in connection with filing problems faced by Scotland Yard officials on the classification and identification of criminals Its appearance then was as a research technique rather than as a mathematization of a psychological model At that time, Galton [1888], who was developing his ideas on correlation, became interested in the classification problem He pointed out that the 12 Bertillon measures to be used for classification were not independent and proposed that the observed measurements be transformed into a set of independent measures He also suggested the method of transformation which is now known as simple or unweighted summation in factor analysis terminology

The use of factor analysis as a research technique in classification problems is still quite current Nevertheless its most interesting aspects are found in its attempts to provide frameworks for the measurement of mental ability These frameworks are conceptual or latent universes which command some psychological or physiological relevance and are then mathematized These mathematizations or mathematical models of factor analysis are the target of this article Naturally they began, as in other disciplines, by attempts at description and explanation of data In this case, the data motivating the models resulted from tests of mental ability Such a pro-

cedure could evoke some skepticism since the existence of tests of mental ability implies a measurement structure in which the tests are imbedded However, sciences usually do not initially grow from sophisticated and developed concepts but rather from experimental data or from dreams that lead to experimental data Sophistication, in fact, usually implies a long period of analysis and study In the usual factor analysis situation, the manifest data (responses to tests of mental ability) are organized in a correlation matrix and conjectures, guesses, hunches, about the latent universe operating to produce the manifest data are usually mathematized These mathematizations are our mathematical models in factor analysis The earlier models by Spearman and Thomson could serve as potential structures for relating the operation of the brain and nervous system to responses to tests of mental ability, the later models of Thurstone and Hotelling serve more as aids in the construction of batteries of tests Holzinger's model, while chronologically later than Thurstone and Hotelling, stands in structure between these two groups On the other hand, in Guttman's model which is the latest study, a point can be made for its serving both of the aforementioned structures

A plunge into factor analysis models can immediately demonstrate the same issues and questions related to model building in any of the sciences Along these lines three, but not necessarily all, kinds of approaches toward conceptualization of a model will be noticed first, of course, is the adventurer who from the data notices orderings, arrangements, patterns, etc which usually permit a more parsimonious accounting of the data, second is the negativist, that is, one who admits that the adventurer has scored a point but then demonstrates that a different mathematization, sometimes a most irrelevant one, will also account for the data, third, but not necessarily last, is one who seizes on a certain mathematization, operates on it in a purely mathematical way and then notices that the newly transformed model can both account for the data and lend itself to reasonable explanation The ordering of these three approaches is in no way intended to show the relative validity of the models produced While all these moves produce a certain amount of intellectual activity and yield models, we are still left with the question of validity of the model even though reproducibility of data may be ascertained This is certainly true for factor analysis In addition, for any one general model there will still be questions of whether certain kinds of observed data can determine the model uniquely These questions of identification are common to model building in general and have been seriously investigated by econometricians and statisticians A detailed account of

these problems and their resolution in factor analysis for the multiple common factor model is given in Anderson and Rubin [1956]

In this paper we will dwell solely on the mathematical models in factor analysis Estimation of parameters in models and distribution theory for these estimates will be mentioned where relevance dictates it, but will only be developed when a central issue or controversy is at stake This will be so when we examine the Hotelling and Thurstone models The same will hold true for factor analysis as a research technique in contrast to factor analysis as a model for mental measurement A search for the most provocative way of presenting the models, their common aspects, and their departures from one another has led to the natural choice of beginning with Spearman's model In fact, the presentation will be chronological with few exceptions although exact chronology is sometimes difficult except for dates of publication

The mathematization of these models was attempted by several scholars of the first half of this century Also to be noted is the almost complete domination of the field by the British and American schools and the cleavage in approach between these two groups in a manner reminiscent of their differences in other sciences Obviously, the emphases, the deletions, and the accounts of the models are the responsibility of the author The author will agree, *a priori*, that there will be sins of omission but trusts that sins of commission will be minimized In addition, no attempts will be made to develop the mathematics or mathematical statistics used throughout the paper

2 THE SPEARMAN MODEL

We enter our story with Spearman's experimental data resulting from tests which by context and expert judgment relate to mental ability Spearman, in his 1904 paper, was concerned both with the measurement of mental ability and with what he considered the futile attempts at intrinsic measurement of mental ability in previous researches Prior studies always related intellective ability to some external criteria such as age or sex and interestingly to sensory discriminatory abilities for such phenomena as sound, light, and weight Until Spearman's paper appeared no definite structure seemed to emerge from the previous measurement studies which incidentally were conducted by the leading psychologists of the day Motivated by the thought that there was a common intellective element in the observed responses to tests of mental ability, Spearman organized some experimentation in a

village school and a preparatory school for boys in England The children were given four achievement tests and two sensory discrimination tests The test data were summarized by correlation coefficients between all pairs of tests and then presented in the form of a correlation matrix *in which* each element, except for *those in the main* diagonal, measures the relationship between a pair of tests

It is instructive to provide the correlation matrix which led to the birth of factor analysis The correlations are based on test scores for 36 boys ranging in age from 9 years, 5 months to 13 years, 7 months

	Classics	French	English	Mathematics	Discrimination of Tones	Musical Talent
Classics		83	78	70	66	63
French	83		67	67	65	.57
English	78	67		64	.54	51
Mathematics	70	67	64		45	.51
Discrimination of Tones	66	65	.54	45		40
Musical Talent	63	.57	51	51	40	

In Spearman's fundamental paper *this collection of correlation* coefficients is examined in some detail The correlation coefficients are presented in matrix form where r_{ij}, the element in the i^{th} row and j^{th} column of the matrix, is the correlation between test i and test j $(i, j = 1, 2, \quad , n)$ with n the number of tests We obtain in general

$$R_{ij} = \begin{bmatrix} & r_{12} & r_{13} & r_{14} & \cdots & r_{1n} \\ r_{21} & & r_{23} & r_{24} & \cdots & r_{2n} \\ \vdots & & & \ddots & & \vdots \\ r_{n1} & r_{n2} & r_{n3} & r_{n4} & \cdots & \end{bmatrix} \qquad (2\ 1)$$

Of course one might question at *this early point in our development* why the correlation matrix is the fundamental summary of the data, and why the diagonal elements are not included especially when from a strictly statistical point of view they should all be equal to one The former query may be handled by the traditional assumption of a normal distribution for the variables which leads directly to the statement that *knowledge of the second order moments* (variances and covariances) uniquely determines all the in-

formation in the data, in addition, correlations rather than covariances are employed because the variables used are usually not commensurate and require the dimensionless correlation coefficient If all the variables are commensurate, factor analysis can be developed from covariances rather than from the traditional correlations The latter query on vacant diagonal cells can be resolved by another assumption of classical factor analysis This states that an observed variable is a linear function of independent common factor variables plus a variable specific only to the observation and independent of all other specific variables and the common factor variables The correlation of the observed variable with itself, namely the variance which is standardized and thus equal to one, then, is made up of two components, the variance due to the common factors which is called the *communality* and the variance due to the specific variable Since each element in the correlation matrix is intended to represent the degree of overlap due to common causes or common factors, the elements in the main diagonal are the communalities and can be less than one since the variance due to the specific variable need not be included Since the communalities are usually unknown, although estimates can be obtained from the data, the main diagonal cells are left vacant at this point

After examining the correlation matrix, Spearman perceived a functional relationship in the data which he labelled 'hierarchical order,' namely for any two columns in the matrix the ratio of any corresponding pair of elements was approximately constant, for example, consider the a^{th} and b^{th} row for columns c and d, then

$$\frac{r_{ac}}{r_{ad}} = \frac{r_{bc}}{r_{bd}}$$

Another manifestation of hierarchical order was that the columns could always be arranged so that the numbers steadily decreased from the upper left hand corner of the matrix to the lower right-hand corner This type of arrangement led Guttman many years later to another factor analysis model which is discussed in Section 7 Naturally, perfect hierarchical order did not exist in Spearman's data but what is relevant and surprising is the small deviation from this arrangement that he encountered for such a small sample size as 36 observations

In general, such a matrix is a symmetric square matrix with elements in the main diagonal not reported Correlation matrices based on tests of mental ability and mental achievement always exhibit positive elements in each

cell including the main diagonal when these become known. We have already explained why these diagonal elements are usually missing. How they are chosen plays an important role in some factor analysis models.

If we look further into the manifestations of hierarchical order in a matrix it can be seen that all two by two determinants, not including the main diagonal, are equal to zero. Thus mathematically a matrix of rank one is almost produced from the matrix array of correlation coefficients resulting from tests of mental ability. In this situation, the elements in the main diagonal can always be chosen to insure that the rank is one and thus the communalities are easily and explicitly determined. Offhand the fact that an $n \times n$ matrix would have rank one simply by chance would and should be regarded with suspicion and cause one to conjecture about the underlying structure that produced this result. Spearman's conjecture, and certainly one which could be considered reasonable, even if later proved inconsistent or untenable (this does not mean that these conclusions are definitive for the situation studied by Spearman), was that a response to a test of mental ability was made up of a general intellective factor and a factor specific to the test, the specific factors being mutually independent of each other and of the general factor. As would happen in most disciplines a linear combination of these two factors was assumed for each test, at least for a first try. If z_j represents the j^{th} test, F the general factor, and S_j the specific factor for the j^{th} test, we have

$$z_j = a_j F + S_j \tag{2.2}$$

where a_j is a parameter of the model which we hope to estimate from data and in a practical way represents the amount of saturation of test j with the general factor. Later a_j will be called the factor loading for test j and will also be viewed as the correlation, r_{z_jF}, between the test and the general factor. For, as is usually done, assume F is normally distributed with mean zero and variance one, in brief, $N(0,1)$ and assume z_j is $N(0,1)$, then S_j is $N(0,\sigma_j^2)$, and $a_j^2+\sigma_j^2 = 1$. Now multiply both sides of (2.2) by F and take the expected value of both sides, we get

$$Ez_jF = a_jEF^2 + ES_jF \tag{2.3}$$

$$r_{z_jF} = a_j$$

Likewise

$$Ez_j^2 = a_jEz_jF + ES_jz_j \tag{2.4}$$

$$r_{jj} = a_j^2 + \sigma_j^2 = 1$$

While r_{jj} as given in (2 4) should ordinarily appear in the main diagonal of the r_{ij} matrix, it is only the a_j^2 component which is the communality as we shall indicate later Now to return to the main argument, namely, does this conjecture at least give rise to the pattern noticed in the observed data, *i e*, hierarchical order or a correlation matrix of rank one?

Consider the correlation r_{ij} produced by the model under the previous definitions and assumptions noted

$$r_{ij} = Ez_iz_j = a_ia_jEF^2 + a_iEFS_j + a_jEFS_i + ES_iS_j \tag{2 5}$$

All terms on the right-hand side of (2 5), except the first, vanish because of mutual independence of specific factors and the general factor Thus

$$r_{ij} = a_ia_j \tag{2 6}$$

if the linear model holds Any two by two determinant of the correlation matrix can now be written as

$$r_{ij}r_{lk} - r_{ik}r_{lj} \tag{2 7}$$

This value is also called the tetrad difference

Under the model just constructed this is equivalent to

$$a_ia_ja_la_k - a_ia_ka_la_j = 0 \tag{2 8}$$

Under the model, the correlation matrix is of rank one provided the diagonal elements are appropriately chosen and thus a latent universe which could produce the manifest data is provided

Another way of mathematizing this particular structure under the normal distribution assumptions just stipulated can be given as follows If F is a factor common to all tests and independent of the mutually exclusive specific factors, then the partial correlation between any two tests obtained by "partialing out" the common factor F must be zero since F, by the model, is the only possible means of providing overlap between the two tests Setting the partial correlation $r_{z_iz_j \cdot F} = 0$, we obtain

$$r_{z_iz_j} - r_{z_iF}\, r_{z_jF} = 0$$

$$r_{ij} = a_i a_j, \tag{2 9}$$

$$r_{z_iF} = a_i, r_{z_jF} = a_j$$

since

This correlation argument may be somewhat more instructive than the linear model, although they both express the same structure

One can see at this point that the Spearman model, which over time has borne such names as the Theory of General Ability, the Single Factor Theory, the Two Factor Theory, and reasonable composites of these terms, has at least two points to recommend it First, it serves to represent a possible latent universe yielding the data observed from tests of mental ability, and second, it has the property that it is rather simple in structure Nevertheless, while these are desirable properties, further evidence is necessary before it can be established as the causal mechanism for the phenomenon of hierarchical order exhibited in this data and this evidence must come with the aid of studies outside the field of mathematization Such outside demonstration, for example, might lie in the field of neurological research

3 A "COUNTER-EXAMPLE" TO THE SPEARMAN MODEL

Another event which could strengthen the importance of the Spearman model would be the complete lack, after study, of other reasonable conjectures about latent universes However, Spearman's conjecture was not long without a competitor The initial indication of competition was in the first of a series of papers by Godfrey Thomson [1916] Practically in Thomson's first words, he stated that his objective was to show that the data brought forward by Spearman in favor of the existence of the General Factor were by no means "crucial' By this he meant that the data were not inconsistent with the notion of a common element through all mental tests, but neither were they inconsistent with its non existence Of course, to demonstrate this last phrase, some mathematization would have to be constructed which would reproduce the data yet have no resemblance to the theory of General Ability

Before we discuss this "counter example' let us state some issues Briefly, both Spearman and his opponents agreed that there are Specific Factors peculiar to individual tests, both sides agreed that there are Group Factors which run through some but not all tests The difference results from Spearman's statement that there is a further single factor which runs through all tests, and moreover that by pooling some tests appropriately the Group Factors can soon be eliminated and a point reached where all the correlations are due to the General Factor alone Of course, the basis for Spearman's reasoning is the existence of hierarchical order in the correlation matrix

Some description of the experiment is interesting Suppose first we simulate the Spearman conjecture Consider ten tests where the numerical score for each test is obtained by the throw of dice If we throw an entirely fresh set of dice to represent an individual's score on each of the ten tests, then clearly the factors are entirely specific and there will be no correlation whatever between the scores in the ten tests If some of the dice are red, and these red dice are left lying and counted in to every score, there will be a General Factor The red dice are of course rethrown when we are finding the scores of the next individual, in whom the General Factor may be greater or less Consider now that white dice whose number varies from test to test are added to the red dice and thrown, the white dice representing the Specific Factors in each test The correlations between tests in this situation will form a perfect hierarchy For example the upper hierarchy in Table I could be obtained in this way The numbers given there are the theoretical values calculated by

$$r = \frac{h}{\sqrt{(h+m)\,(h+k)}} \tag{3 1}$$

where $(h + m)$ dice are thrown on the first throw, h left lying, and k thrown to form with h the second throw of $(h + k)$ dice The number of dice common to all the tests is 19, and the total number of dice in each set is shown below Table I

Now note the lower numbers in each cell in the table They do not form a perfect hierarchy but nevertheless a very good approximation with no reversals, that is, the numbers steadily decrease from the upper left-hand corner to the lower right-hand corner This almost perfect hierarchy contains no General Factor whatever and was constructed in the following way We have an arrangement in which ten tests depend on 145 factors of which 109 are quite specific and only occur in one test each, while the remaining 36 are group factors which run through more than one test each, but never through all The distribution of the 36 group factors is shown in Figure 1 Most of them run through only two, three, or four tests, three of them run through five tests, but not one runs through more than five tests out of the ten There is therefore nothing approaching a General Factor The number of purely Specific Factors in each test is also given

To obtain actual scores for an individual, thirty six dice, marked so that each was recognizable, would be thrown, and the score of each placed in the proper place in Figure 1 These are the Group Factors Then in addition the dice representing the Specific Factors would be thrown These would be

entirely separate for each test, for example, fourteen dice would be thrown to complete test *f* The scores of the various tests could then be added up and the totals would be analogous to the scores of a single individual in the ten tests This whole enterprise could then be repeated for each of the individuals

To make this clear, consider two tests in detail, say tests *d* and *e* For test *d*, twenty dice in all have been thrown Of these five are left lying and sixteen other dice thrown to complete test *e* The correlation between test *e* and test *d* will therefore be by the formula already mentioned

$$r = \frac{h}{\sqrt{m+h)\,(h+k)}} = \frac{5}{\sqrt{(20)\,(21)}} = 0\ 244 \qquad (3\ 2)$$

By the same formula the correlations between all pairs of tests can be found They are the lower numbers shown in Table I and form as has been said an excellent hierarchy

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
a	×	×	×	×	×	—	×	×	×	×	×	×	×	×	×	—	—	—
b	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	—	—	—
c	×	×	—	×	—	×	×	—	—	×	×	—	—	×	—	—	—	—
d	×	—	×	—	×	—	—	×	—	—	—	×	—	—	×	×	—	—
e	—	×	—	—	×	—	—	—	×	—	—	—	×	—	—	—	×	×
f	—	—	×	—	—	—	×	—	—	×	—	—	—	—	—	—	×	—
g	—	—	—	×	—	—	—	×	—	—	—	—	×	—	—	—	—	—
h	—	—	—	—	—	×	—	—	—	—	—	×	—	—	—	—	×	—
k	—	—	—	—	—	—	—	—	×	—	—	—	—	×	—	×	—	—
l	—	—	—	—	—	—	—	—	—	—	×	—	—	—	×	—	—	×

	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	S
a	—	—	—	—	×	—	—	×	—	×	×	×	—	×	×	×	×	×	0
b	—	—	—	—	—	—	—	—	—	—	—	—	—	×	×	×	×	×	0
c	—	—	—	—	—	×	×	×	—	×	×	×	—	×	×	×	×	×	1
d	—	—	—	—	—	×	×	×	×	×	—	—	—	×	×	×	×	×	3
e	—	—	—	—	—	×	—	—	—	—	×	—	×	×	×	×	—	—	9
f	×	×	—	—	—	—	—	—	×	—	×	—	—	—	—	—	—	—	14
g	×	—	×	×	—	—	—	×	—	—	—	—	×	—	—	—	—	—	16
h	—	—	×	—	×	—	—	—	—	×	—	—	—	—	—	—	—	—	20
k	—	×	×	—	—	—	—	—	—	—	—	×	—	—	—	—	—	—	22
l	—	×	—	×	×	—	—	—	—	—	—	—	—	—	—	—	—	—	24

Fig 1 Distribution of *Group Factors**

a, b, c, etc are the names of tests '
1, 2, 3, etc are factors '
For example, factor number **15** *(perchance* visual memory) runs through tests **a, b, d, l.**
In addition to the Group Factors there are Specific Factors the number of which in each test is indicated under **S**

(*) Figure reprinted from *British Journal of Psychology*, Vol VIII, 1916 p. 277

The main point is that while the upper and lower correlations in each cell in Table I are theoretical values for the two different simulations, the differences in each cell, if based on a sample of 36 individuals as in Spearman's experiment, are so small that by the criteria laid down by Spearman for hierarchical order both would have to be accepted. Actually, Thomson did use the same mock-up for overlapping group factors and obtained an empirical correlation matrix for 36 individuals. To do this, 5,220 dice were thrown, in 36 groups of 145 each, to represent ten tests for a class of 36 individuals. These data also permitted acceptance of hierarchical order.

A complete discussion of the details of the experiment seems warranted for this essay since it permits us to belabor the following major point at this juncture, namely that Thomson has been essentially the negativist because he has demonstrated that hierarchical order alone cannot permit discrimination between a General Factor, and overlapping Group Factors without the presence of a General Factor. Moreover, since the essence of this article is in the initiation and development of mathematical models, we will discuss the Spearman-Thomson contributions in some detail.

Thomson's counter-example, published in 1916 but prepared in 1914, initiated an active and bitter controversy which was carried mainly in British journals by the two protagonists and their supporters and lasted into the late nineteen-thirties. Since British journals are not so circumspect as American journals in reporting controversy, some excerpts may be interesting. Here are some excerpts from Spearman's comments which appeared right after the ending of Thomson's first paper. It should be noted that Spearman was engaged in war service at the time.

> At the present time, I have little leisure, or indeed inclination, for controversy. But the foregoing clever and interesting paper seems likely to produce grave misunderstanding unless some brief comments are attached.
>
> This brings us to the real weakness in the paper, one which, frankly, I should not have expected from the author.
>
> This absence of any stated principle in the distribution of factors is quite sufficient to invalidate all the conclusions of the foregoing paper I may, however, mention that I propose in a subsequent paper to demonstrate the nature of the principle really involved. It happens to have long been known to me, being an integral part of the proof of the above mentioned effect of random overlapping On this principle being established, the author's method will at length be able to bear the good fruit which its ingeniousness deserves It will be found to have the same curious issue as the other evidence brought against the 'theory of two factors', for it will turn out to be really a witness in favour.

Spearman's argument was that Thomson had arduously sought and found a "special arrangement" whereby hierarchical order could be obtained

without the existence of a General Factor, that this could be obtained by chance from a simulated setup only an infinitesimal amount of times, that even if these arrangements arose they might involve psychological absurdities In fact, if we refer back to his direct words in the previous paragraphs, Spearman is prepared to show that Thomson's demonstration actually re-introduces the General Factor because the counter example is allegedly such a rare occurrence

In a subsequent article, Thomson [1919] replied to these comments but even more important performed the positive act of framing a new theory of mental ability First, a few words about Thomson's rebuttal on the question of using a "special arrangement" as the counter example to the theory of General Ability In the first place, any illustration which provides hierarchical order without a General Factor should cause some concern However, using the same argument, one can say that the probability of any given general factor occurring from Thomson's artificial experiment just by chance also is extremely small, for this can occur only when the group factors operate so that a prudent pooling of tests eliminate the overlap arrangement Moreover, many more arrangements lacking the general factor but yielding a fair semblance of hierarchical order can be produced by some shuffling in Thomson's dice experiment Given his dice experiment, Thomson stated that the number of hierarchy producing arrangements immediately deducible from his original "special arrangement" exceeds the total possible cases due to a general factor at least 10^{69} times Thus excluding all cases of a *real* General Factor, there will be millions and millions of hierarchy-producing arrangements not containing the ghost of a general factor and on the evidence available the existence of such a factor is extraordinarily improbable

As stated in the previous paragraph, Thomson then formulated a new theory of ability which leads to hierarchical order No doubt the probing for other causes of hierarchical order, even artificially mathematized situations, led to Thomson's new formulation which now could also be expressed in the form of a theory with psychological relevance

4 THOMSON'S THEORY

The first statement of Thomson's theory [1919] by its originator is as follows

The mind, in carrying out any activity such as a mental test, has two levels at which it can operate The elements of activity at the lower level are entirely specific,

but those at the higher level are such that they may come into play in different activities Any activity is a *sample* of these elements The elements are assumed to be additive like dice, and each to act on the 'all or none' principle, not being in fact further divisible

Before going into these principles in more detail, suffice it to say that the controversy over the latent universe producing hierarchical order continued for many years Spearman maintained the General Factor Theory while Thomson agreed that it certainly could produce hierarchical order but that more than likely his theory, which he called the 'Sampling of the Bonds Theory," might be a better, or at least an alternate, explanation

In a previous section we remarked on the tendency to low rank in correlation matrices based on tests of mental ability which is really the foundation on which Spearman's theory rests Actually the proponents of the theory of General Ability were also ignoring a host of large specific factors in achieving this low rank and concluding that there must be something in it Let us repeat that an n by n matrix of rank one is remarkable and deserves the study it has had for many years Thomson also agreed that this phenomenon was remarkable but presented an almost opposite explanation for it He felt that instead of showing that the mind has a definite structure, the low rank demonstrated, on the contrary, that the mind has hardly any structure In fact, if the rank in all cases was equal to one, then the mind had no structure at all but was completely undifferentiated Thomson claimed that it was the departures from rank one which indicated structure and that a significant fact to be culled from experimental reports was that batteries of tests lead to higher rank in correlation matrices for adults than for children Probably because in adults such forces as education and vocation have imposed a structure on the mind which is absent in children With regard to this point, Spearman also made some interesting comments in a footnote in his fundamental paper [1904, p 276] This arose in connection with a comparison he made to his data of some relevant data collected at Columbia University around the turn of the century by Cottell and Farrand and then published in 1901 by Wissler (*) The Columbia data met the test of hierarchical order only in a limited way Here are Spearman's comments

> But a university is clearly not the place in which to look for natural correspondence between functions, at that time of life, strong ties of a wholly artificial sort have intervened, each student singles out for himself that particular group of studies tending to his main purpose and devotes to them the most judicious amounts of relative energy To determine natural correlations, we must go to where the pupils meet each other in every department on relatively equal terms

(*) *Psychological Review* Monograph Supplement, June 1901

Both Thomson and Spearman seem to be in agreement on this possibility leading to correlation matrices of rank higher than one

What Thomson meant by the absence of structure is the absence of any fixed or strong linkages among the elements of the mind so that a 'test" *samples* whatever of those *total* elements or components can be assembled in the activity under consideration Even assuming the mind is an unstructured continuum, the word "element" or "component" can still have some meaning as an atomistic part of the mental ability continuum but there is no commitment to this conjecture For concreteness, Thomson found it convenient to identify these elements, on the mental side, with something of the nature of Thorndike's "bonds," and on the bodily side with neurone arcs This is how the word "bond" entered into Thomson's Sampling Theory All a "bond" means is some very simple unit of the causal background Some may be inherited, some may be due to education, some possibly to other causes

Suppose now that we have a causal background which comprises innumerable bonds and any measurement we make is a function of a sampling of that background, all samples being possible If we obtain a number of different measurements from this assumption and find their intercorrelations, the matrix of these correlations will tend to be hierarchical, or at least tend to have a low rank Now all this need have nothing to do with the mind but can arise simply as a mathematical necessity from the assumptions imposed, whatever material is used to illustrate it However, the fact that this can serve as a causal background for the operation of the mind and that as a result low rank will occur joins together mathematical and psychological relevance In fact, there now seems to be neurological evidence that the brain is more an undifferentiated mass of neurone arcs than an especially compartmentalized organ—a situation which could be relevant for the latent factor models we will discuss in later sections

Now let us observe how the Sampling of the Bonds Theory produces a tendency to zero tetrad differences, that is, correlation matrices of rank one As a simple illustration, Thomson [1927b] proposed a "mind" which can form only six bonds This mind is submitted to four "tests" which are of different degrees of richness, one requiring the joint action of five bonds, the other of four, three, and two, respectively These four tests when we give them to a number of such minds, yield correlations with one another

We have only specified the richness of each test, but have not said which bonds for each ability There may, therefore, be different degrees of overlap between them, though some will be more frequent than others if we form all

the possible sets of four tests which are of richness five, four, three, two. If we call the bonds a, b, c, d, e, and f, then one possible pattern of overlap could be the following:

Test	Bonds					
1	a	b	c	d	e	
2		b	c	d	e	
3				d	e	f
4			c	d		

(4.1)

For simplicity, suppose these bonds to be equally important, and use the natural formula

$$\text{Correlation} = \frac{\text{overlap}}{\text{geometric mean of the two totals}} \tag{4.2}$$

The correlation matrix is then computed yielding

	1	2	3	4
1		$4/\sqrt{20}$	$2/\sqrt{15}$	$2/\sqrt{10}$
2	$4/\sqrt{20}$		$2/\sqrt{12}$	$2/\sqrt{8}$
3	$2/\sqrt{15}$	$2/\sqrt{12}$		$1/\sqrt{6}$
4	$2/\sqrt{10}$	$2/\sqrt{8}$	$1/\sqrt{6}$	

(4.3)

and we notice that in this particular pattern all three tetrad-differences are zero. However, if we picked our four tests at random, insuring the same four degrees of richness, we would not always get the same pattern, in fact, we would get it only 12 times in 450. Nevertheless it is one of the most probable patterns. In all, 78 different patterns of the bonds are possible for the "richness" situation depicted above, the probability of each pattern ranging from 12 in 450 down to 1 in 450. We can calculate the tetrad-differences for each of the 78 possible patterns of overlap which can occur. There will be 450 values of tetrad-differences distributed around a mean of zero with variance equal to 0.04.

Thus if there are no linkages among the bonds, and if all possible samplings of the bonds are taken, the average of all the tetrad differences will be zero. As the number of bonds in the mind increases, the tetrad differences crowd closer and closer to zero. For example, assume the above experiment is conducted in a universe of men whose minds could form twelve bonds, the four tests requiring ten, eight, six, and four of these bonds. The work of calculating all the possible patterns of overlap, and the frequency of each, becomes enormous. Thomson [1927b] demonstrated that there would be 1,257 differ-

ent correlation matrices. All in all there will be 1,087,110 instances required to represent the distribution of tetrad differences. Thomson made all calculations and presented the frequency distribution. The average tetrad-difference was again exactly equal to zero but the variance is now 0.018. Thus doubling the number of bonds in the mind has practically halved the variance of the tetrad-difference. This indicates that increasing the number of potential bonds supposed to exist in the mind to anything like what must be its true figure would give variances equal to zero for all practical purposes, and thus provide a value of zero (not an average value of zero) for tetrad differences. In other words, we are back to hierarchical order and zero tetrad-differences but from a completely different causal background than the theory of General Ability. In fact, it is now difficult to see how tendency toward hierarchical order can be the sole foundation for any special theory of ability.

Suppose we assume in general that the number of bonds is N, and the four test variates contain p_1N, p_2N, p_3N, p_4N of the whole number of bonds where $0 < p_i < 1$ and their sum is one. Upon these assumptions, Thomson [1927a] showed that the most probable value of a tetrad-difference is zero. Mackie [1929] showed that the mean value of the tetrad-difference is zero and that the variance σ^2_{TD} is

$$\sigma^2_{TD} = \frac{1}{N-1}\Big\{p_1p_3+p_2p_4+p_1p_4+p_2p_3-2(p_1p_2p_3+p_1p_2p_4+p_1p_3p_4+p_2p_3p_4)+4p_1p_2p_3p_4+\frac{2(N-2)}{(N-1)^2}(1-p_1)(1-p_2)(1-p_3)(1-p_4)\Big\} \quad (4.4)$$

Here we can see how the variance vanishes as N increases as Thomson predicted from his empirical calculations with $N = 6$ and 12. In fact for large N, the variance decreases as $1/(N-1)$. Even for $N = 6$ and $N = 12$ we see that $0.018 = \frac{6-1}{12-1}(0.40)$ where 0.018 was the computed variance for $N = 12$, 0.40 the computed variance for $N = 6$.

5. RECONCILIATION OF THE SPEARMAN AND THOMSON MODELS

Since the Spearman and Thomson models both lead to zero tetrad-differences it should be possible to demonstrate how one can go from the mathematization of one model to the mathematization of the other. Thomson [1935] did show how to form the transformation matrix which turns the

Spearman equations into equations which represent the Sampling Theory. For simplicity, he wrote out the case of three test variates in full but the method is perfectly general. Consider the general Spearman equations

$$y = Af \tag{5.1}$$

as referring to three variates only

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_1 & s_1 & & \\ a_2 & & s_2 & \\ a_3 & & & s_3 \end{bmatrix} \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix} \tag{5.2}$$

where the multiplication of the matrices is carried out as usual, row by column, and f_0 represents the general ability.

Consider now the orthogonal matrix

$$\begin{array}{c|ccc|ccc|c}
a_1a_2a_3 & s_1a_2a_3 & a_1s_2a_3 & a_1a_2s_3 & s_1s_2a_3 & s_1a_2s_3 & a_1s_2s_3 & s_1s_2s_3 \\
\hline
s_1a_2a_3 & -a_1a_2a_3 & s_1s_2a_3 & s_1a_2s_3 & -a_1s_2a_3 & -a_1a_2s_3 & s_1s_2s_3 & -a_1s_2s_3 \\
a_1s_2a_3 & s_1s_2a_3 & -a_1a_2a_3 & a_1s_2s_3 & -s_1a_2a_3 & s_1s_2s_3 & -a_1a_2s_3 & -s_1a_2s_3 \\
a_1a_2s_3 & s_1a_2s_3 & a_1s_2s_3 & -a_1a_2a_3 & s_1s_2s_3 & -s_1a_2a_3 & -a_1s_2a_3 & -s_1s_2a_3 \\
\hline
s_1s_2a_3 & -a_1s_2a_3 & -s_1a_2a_3 & s_1s_2s_3 & a_1a_2a_3 & -a_1s_2s_3 & -s_1a_2s_3 & a_1a_2s_3 \\
s_1a_2s_3 & -a_1a_2s_3 & s_1s_2s_3 & -s_1a_2a_3 & -a_1s_2s_3 & a_1a_2a_3 & -s_1s_2a_3 & a_1s_2a_3 \\
a_1s_2s_3 & s_1s_2s_3 & -a_1a_2s_3 & -a_1s_2a_3 & -s_1a_2s_3 & -s_1s_2a_3 & a_1a_2a_3 & s_1a_2a_3 \\
\hline
s_1s_2s_3 & -a_1s_2s_3 & -s_1a_2s_3 & -s_1s_2a_3 & a_1a_2s_3 & a_1s_2a_3 & s_1a_2a_3 & -a_1a_2a_3
\end{array} \tag{5.3}$$

This matrix is symmetrical about the principal diagonal and it has other symmetrical properties which are best seen if it is studied after dissection, as indicated by the lines. If now the column vector

$$f = \begin{bmatrix} f_0 \\ f_1 \\ f_2 \\ f_3 \end{bmatrix} \tag{5.4}$$

is replaced by

$$f = Tz, \tag{5.5}$$

where T is the upper four rows of the above orthogonal matrix, and z is a column vector of eight elements ($z_1, z_2, \ldots, z_8$) (since T has eight columns), we get, by straightforward multiplication of AT in the equation

$$y = ATz, \tag{5.6}$$

the result

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_2a_3 & 0 & s_2a_3 & a_2s_3 & 0 & 0 & s_2s_3 \\ a_1a_3 & s_1a_3 & 0 & a_1s_3 & 0 & s_1s_3 & 0 \\ a_1a_2 & s_1a_2 & a_1s_2 & 0 & s_1s_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \\ z_7 \\ z_8 \end{bmatrix}, \tag{5 7}$$

or written as ordinary equations

$$y_1 = a_2a_3z_1 + s_2a_3z_3 + a_2s_3z_4 + s_2s_3z_7$$
$$y_2 = a_1a_3z_1 + s_1a_3z_2 + a_1s_3z_4 + s_1s_3z_6$$
$$y_3 = a_1a_2z_1 + s_1a_2z_2 + a_1s_2z_3 + s_1s_2z_5$$

Notice the important phenomenon, at this time, that z_8 has disappeared entirely What it is will become apparent shortly

Thus far these equations will give the same correlations as the Spearman equations with which we started and therefore, in the general case of many variates, instead of three, will give zero tetrad-differences Remember they are only one among the infinity of sets into which the Spearman equations can be transformed However, we can give this particular set an interpretation in terms of Thomson's Sampling Theory At present, each y is in standard form However, multiply the equation for y by a_1, that for y_2 by a_2 etc so that they become

$$a_1y_1 = a_1a_2a_3z_1 + a_1s_2a_3z_3 + a_1a_2s_3z_4 + a_1s_2s_3z_7$$
$$a_2y_2 = a_1a_2a_3z_1 + s_1a_2a_3z_2 + a_1a_2s_3z_4 + s_1a_2s_3z_6$$
$$a_3y_3 = a_1a_2a_3z_1 + s_1a_2a_3z_2 + a_1s_2a_3z_3 + s_1s_2a_3z_5$$

The three variates a_1y_1, a_2y_2, a_3y_3 could now have the following explanation Each is composed of $a_i^2\mathcal{N}$ ($i = 1,2,3$) small equal components (bonds) drawn from a pool of $\mathcal{N}$ such elementary components (bonds), each component, all or none in response In this situation, the number of elementary components which would most probably occur in each and every variate is $a_1^2a_2^2a_3^2\mathcal{N}$ and this block of elementary components forms a general factor (not Spearman's general ability factor) with standard deviation, over the population of all possible samplings of the elementary components, proportional to $a_1a_2a_3$ This is represented by $a_1a_2a_3z_1$ in the equations. The number of

elementary components likely to be found in variates 1 and 2 but not in variate 3 will be $a_1^2a_2^2s_3^2N$ ($s_3^2=1-a_3^2$), and these are represented by $a_1a_2s_3z_4$ which in the above equations occurs in the first two variates but not in the third, etc The z_8 which has disappeared from the equations represents the elementary components, most probably $s_1^2s_2^2s_3^2N$ in number, which are not drawn from the pool in any one of the three drawings

Thus, as was to be expected and as Thomson ingeniously showed, the Spearman model and the Thomson model, both of which rest on zero tetrad-differences, can be obtained one from the other by strict mathematical operations Enough has probably been said in this paper to indicate Thomson's summations which were made in several papers and over a period of many years Spearman, also, in several papers over a period of some thirty years, continued to present his rationale for the theory of General Ability However, much of Spearman's discussion and counter arguments have lost their immediate or substantive relevancy for the focus of this essay and have not been given here

6 ALTERNATIVES TO HIERARCHICAL ORDER

The Spearman and Thomson models and discussions of these models take us in time from 1904 to the latter half of the nineteen thirties The more fashionable and currently used multiple factor models began to appear on the scene in the early part of the nineteen thirties These can be looked upon as natural extensions of the Spearman model We will discuss these models in short order However, at this point we will depart from a strictly chronological development and next consider a model by Louis Guttman [1954] which appeared in the early nineteen-fifties and which can in a specific way also be viewed as a natural extension of Spearman's Theory of General Ability

Before we introduce Guttman's model, let us reconsider the issues by examining why extensions of the Spearman model or Thomson model are necessary, and if they are, how such extensions might develop Both the Theory of General Ability and the Sampling of the Bonds Theory rest on zero tetrad differences or equivalent correlation matrices of rank one When correlation matrices depart from rank one, either theory becomes suspect or it at least must be qualified in some way Those who are trying to maintain the Spearman model can argue as some did that when this situation of higher ranks occurred, the tests of mental ability were not selected "correctly" and that care in this matter could leave the theory of General

Ability inviolate In spite of repeated evidence of lack of rank one in the correlation matrices formed from tests of mental ability, the existence of a General Ability factor need not be denied by Spearman supporters but some qualifications in the remainder of his conjectures would be essential Qualifications which contain some psychological relevance and which could also lead to correlation matrices with higher ranks would be in order In these contexts we will later review the Holzinger, Hotelling, and Thurstone models which successfully mathematized qualifications in the Spearman model and produced correlation matrices of rank higher than one

It is difficult to see what this kind of empirical evidence could mean for qualifications in the Thomson model Certainly the mathematization of a modified Thomson model which would have psychological relevance and produce correlation matrices with rank higher than one has not been attempted successfully However, we have already discussed what this manifestation of higher ranks could mean to Thomson, namely, the effect of education and vocational training as forces on the highly complex but undifferentiated mass known as the brain Certainly these forces could nullify the assumptions of Thompson's theory and the consequent manifestation of zero tetrad-differences As previously mentioned, Thomson considered the differences in rank of correlation matrices observed from experimentation with children and with adults (children, tendency toward rank one, adults, tendency toward higher rank) as highly significant for an appraisal of his conjectures Obviously some restrictions on how the bonds are linked or how the sampling operates can be mathematized and this can mathematically lead to higher ranks This approach has not yet been investigated

The Guttman model, which is the latest in time, is now interposed before the other models because it begins with a re examination of the pattern of hierarchical order first observed by Spearman, and in some of its aspects rests on zero tetrad differences, albeit in a restricted way Actually the term hierarchical order was dispensed with and forgotten in the work of the other model builders whom we shall discuss after Guttman

Guttman called his approach to factor analysis the Radex model and we shall see as the discussion unfolds how this apt name is chosen. The Radex has its roots in Guttman's work on scaling theory Because of this, the notion of order plays an important role in his factor structure This should be contrasted with other approaches which, except for Thomson's, are based on common factors where *order* of factor or test plays no role The Radex model is a definite contribution as we shall soon see, and does relate to other models, but some caution in the claims advertised by Guttman are in order, especially

the unifying framework the model allegedly provides for opposing schools of thought in factor analysis The following quote from Guttman's original paper is illustrative

My own first attempt at an alternative theory to Spearman's was a mathematical success, but an empirical failure It had beautiful mathematical properties not possessed by any common factor theory, but I could find no data for which the mathematics was appropriate in practice More recently, I have learned how to modify my theory, and correspondingly have found it actually useful for certain empirical data This theory I call that of the *simplex* A related theory I call that of the *circumplex* The more comprehensive theory of the *radex* (which combines simplexes and circumplexes simultaneously) has properties which severally have been emphasized or sought for by one or more of the existing common-factor schools In a sense, the new approach unifies all the older approaches simultaneously Each older approach will find its main point more or less justified by the radex, and not at all in contradiction to the main points of other schools On the other hand, this new approach does away with some of the preconceptions of the older approaches, only by so doing could the older theories be unified

7 THE GUTTMAN MODELS

We now embark on Guttman's analysis, for its contribution and relevance are not diminished by any grandiose claims its author may have for it In its simplest form the Radex states that responses to tests of mental ability are controlled by two dimensions, one measuring complexity of test, and the second measuring kind of test In Guttman's mapping of abilities, distance from an origin indicates complexity of test and direction from the horizontal component through the origin indicates subject matter of the test Obviously *order* in complexity of test and even in kind of test will now play a crucial role whereas it did not in our previously discussed theories Another change is that we are now saying that a response to a test of mental ability is characterized by two dimensions whereas in Spearman's characterization only one dimension was apparently necessary and Thomson's model cannot receive an immediate contrast. However, Spearman's model essentially contained two factors for every test response the general ability factor which could be likened to Guttman's dimension of complexity and the specific factor which could be likened to kind of test or subject matter of test. It is also possible to conceive of and study higher-dimensional radexes

It is easiest to digest the Radex model if we analyze each of the two di-

mensions independently Also we shall only develop the simplest structures for each dimension and for their combination Suppose now we consider all tests of the same kind, say, of arithmetic ability Differences in these tests exist largely through a degree of complexity Consider, for example, tests in addition, subtraction, multiplication, and division Guttman called such a set of variables a Simplex, because they possess a simple order of complexity since they can be arranged in a simple rank order from least complex to most complex

Likewise, all tests of the same degree of complexity will differ among themselves in the kind of ability they define but while there may be *order* here it is not in a "least" to "most" sense Since this order has no beginning or end, Guttman conceived that from an elementary standpoint this could defined by a circular order and called a set of variables obeying such a law a Circumplex, to designate a circular ordering By this, Guttman meant that there is an order in the placement of tests of equal complexity around a circle which is unique and meaningful

In the more general case, tests can differ among themselves simultaneously both in degree and in kind of complexity This general two-dimensional structure is the Radex Guttman's motive is to present, on the basis of empirical work, a detailed radex map of some human abilities. Naturally this assumes that the data will not prove to be inconsistent with the Radex model. Actually, Guttman's Factor Theory does reproduce some observed data quite well so that questions of validity rather than questions of tenability will probably be uppermost in the future

Suppose we now return to the position described before, namely, that Spearman's theory of General Ability or a single common factor hypothesis is negated Guttman's Perfect Simplex presents an alternative kind of single factor which still maintains hierarchical order and does not require exploration of a theory of common factors

Consider n tests $t_1, t_2, \ldots, t_n$ which differ only on a single complexity factor These could be the arithmetic tests we discussed before Assume t_1 is the least complex, t_2 requires everything t_1 does, and more, similarly, t_3 is more complex than t_2, requiring everything t_2 does and more Obviously t_3 is also more complex than t_1 In general then, test t_{j+1} is more complex than t_j and thus requires what all the preceding tests require, plus something more Now let g denote the total complexity factor of which all tests are composed in various degrees We are purposely using the letter g in this context because traditionally it has been used as a label for Spearman's

general ability factor and by design we omitted it from our Spearman discussion Our purpose here is to maintain g as a mathematical artifice rather than a substantive factor for the developments discussed in this paper Guttman's basic hypothesis can now be stated as

$$r_{jg \cdot k} = 0 \qquad (j < k) \qquad (7\ 1)$$

where the left hand side of the equation is the partial correlation coefficient between test j and g, "partialing out" test k which is more complex than test j This partial correlation coefficient must be zero because no overlap is possible in this situation under the stated conditions This is very similar to the basic Spearman hypothesis except that there *order* played no part in the considerations (see page 279) We know from the relationship between a first order partial correlation and zero order partial correlations (the ordinary correlation coefficient) that Guttman's hypothesis leads to

$$r_{jg} = r_{jk}\, r_{kg} \qquad (j < k)$$

or

$$r_{jk} = \frac{r_{jg}}{r_{kg}} \qquad (j < k), \qquad (7\ 2)$$

a formula which is meaningless without an ordering concept For example, r_{jk} cannot exceed one but this condition is not violated since $r_{jg} < r_{kg}$ for $j < k$ The analogous Spearman relationship is

$$r_{jk} = r_{jg}\, r_{kg}, \qquad (7\ 3)$$

or the correlation between two tests, j and k, is the product of the saturations of each test with the general factor (factor loadings) and of course no initial ordering of the tests is necessary As we know, Spearman's hypothesis leads to a gradient in the test correlation matrix that descends as one departs from the upper left corner to the lower right corner of the matrix This was one manifestation of hierarchical order

Let us now observe the correlation matrix for Guttman's hypothesis by taking a special situation where the tests are equally spaced in their complexity, that is, $r_{j,j+1} = r_{k,k+1}$ Using Guttman's illustration for five tests, consider complexity loadings of $(\ 6)^5 = 07776$, $(\ 6)^4 = 1296$, $(6\ )^3 = 216$, $(\ 6)^2 = 36$, and 6, respectively Then the correlation matrix for this hypothetical, equally spaced, perfect simplex is given in Table II

TABLE II (*)

Test Intercorrelations for a Hypothetical, Equally-Spaced, Perfect Simplex

Test	Complexity Loading	t_1 .07776	t_2 .1296	t_3 216	t_4 .36	t_5 6
t_1	.07776	1 0	6	36	216	1296
t_2	.1296	6	1 0	6	36	216
t_3	.216	36	6	1 0	6	36
t_4	36	216	36	6	1 0	6
t_5	.6	1296	216	.36	6	1 0
Total		2 3056	2 7760	2 9200	2 7760	2 3056

This time, the largest correlations are all next to the main diagonal and taper off, as one goes to the upper right and lower left of the table If the tests are not equally spaced in their complexity, then the gradient will not be as sharp but the general character will be maintained A set of tests whose observed intercorrelations satisfy the Guttman partial correlation coefficient condition will be said to form a *perfect simplex*. Notice that here we are saying that mathematically $r_{jk}=\frac{a_j}{a_k}$ $(j \leq k)$ where a_j is a parameter belonging to test j which need be defined only up to a constant of proportionality Consider now any two-by-two determinant of the correlation matrix as long as all four of its elements are on one side of the main diagonal

$$\begin{vmatrix} r_{jk} & r_{j\,k+q} \\ r_{j+p\,k} & r_{j+p,k+q} \end{vmatrix} = \begin{vmatrix} \dfrac{a_j}{a_k} & \dfrac{a_j}{a_{k+q}} \\ \dfrac{a_{j+p}}{a_k} & \dfrac{a_{j+p}}{a_{k+q}} \end{vmatrix} = 0 \qquad \begin{matrix} (j \leq k) \\ \\ (j+p \leq k+q) \end{matrix} \tag{7 4}$$

A determinant of this type which contains elements from both sides of the main diagonal does not, in general, vanish Thus tetrad differences are always zero subject to the conditions given above In Spearman's model, all tetrad-differences were zero without reference to the main diagonal condition because there was no problem of order of complexity among the variables More mathematical development of the matrix Guttman has encount-

(*) Table II reprinted from Louis Guttman, "A New Approach to Factor Analysis the Radex", Chapter 6 in *Mathematical Thinking in the Social Sciences*, edited by Paul F Lazarsfeld, 1954, p 271

ered could be helpful in the development of his Simplex model. Guttman discussed quite a few variations of the Simplex model but we will only touch upon them lightly in subsequent sections.

At this point let us examine Guttman's second dimension, namely direction rather than distance from origin or substantively, kind of test rather than complexity of test. We intentionally have considered only the simplest form of the Simplex and now will consider the simplest form of the Circumplex model, the circular order law. When the less simple models arise solely through complication in a fundamental simple law and do not have some external significance, less simple models will not be discussed.

Let us begin with a very special example of a Circumplex. Once again using Guttman's illustration, we have five tests $t_1, t_2, \ldots, t_5$ which have a circular order determined by five elementary components c_1, c_2, c_3, c_4, c_5. We shall call a circumplex *uniform* if each test is a function of an equal number m of the n elementary components. If our $n = 5$ tests form a uniform circumplex with $m = 3$, and if the components are additive, the structure could be written as follows

$$\begin{aligned}
t_{1i} &= c_{1i} + c_{2i} + c_{3i} \\
t_{2i} &= \phantom{c_{1i} +{}} c_{2i} + c_{3i} + c_{4i} \\
t_{3i} &= \phantom{c_{1i} + c_{2i} +{}} c_{3i} + c_{4i} + c_{5i} \\
t_{4i} &= c_{1i} \phantom{+ c_{2i} + c_{3i}} + c_{4i} + c_{5i} \\
t_{5i} &= c_{1i} + c_{2i} \phantom{+ c_{3i} + c_{4i}} + c_{5i}.
\end{aligned} \tag{7.5}$$

Notice how closely this mathematization represents one pattern in Thomson's Sampling Theory for five tests applied to a mind with five bonds, each test having a "richness" of three, that is, each test response is a sampling of three of the five bonds.

In general, for a uniform circumplex we can write

$$t_{ji} = \begin{cases} c_{ji} + c_{j+1,i} + \cdots + c_{j+m-1,i} & (j \leq n-m+1) \\ c_{1i} + c_{2i} + \cdots + c_{j-n+m-1,i} & (j > n-m+1). \end{cases} \tag{7.6}$$

Given a general uniform additive circumplex and assuming all elementary components to be uncorrelated, we obtain

$$r_{c_p c_q} = 0 \quad (p \neq q). \tag{7.7}$$

Suppose we also assume all elementary components have equal variances, say,

$$\sigma^2_{\epsilon_1} = \sigma^2_{\epsilon_2} = \qquad = \sigma^2_{\epsilon_n} = \sigma^2 \tag{7 8}$$

and $m > n/2$, then the correlation, r_{jk}, between tests j and k which differ only in kind and not in complexity is

$$r_{jk} = \begin{cases} 1 - \dfrac{k-j}{m} & 0 \leq k-j < n-m \\ 1 - \dfrac{n-k+j}{m} & n-m \leq k-j < n \end{cases} \tag{7 9}$$

The restriction that the number of elementary components in each test be more than half the total number of components is necessary to limit the number of zero correlations between tests, since in practice a large bloc of zero correlations is not anticipated

Again referring to a Guttman illustration of a numerical example, consider the case where $n = 6$ and $m = 4$ We get the following test correlation matrix in Table III

TABLE III(*)

The Intercorrelations for an Equally Spaced, Uniform, Perfect, Additive Circumplex When $n = 6$ and $m = 4$

Test	t_1	t_2	t_3	t_4	t_5	t_6
t_1	1 00	75	50	25	.50	75
t_2	75	1 00	75	50	25	.50
t_3	50	75	1 00	75	50	.25
t_4	25	50	75	1 00	75	50
t_5	50	25	50	75	1 00	75
t_6	75	50	25	50	75	1 00
Total	3 75	3 75	3 75	3 75	3 75	3 75

Notice that each row of the table has the same entries as the preceding row except that the entries are moved one space to the right, the end entry in one row moving to the beginning of the next row This type of matrix is called a circulant It also appears in serial correlation studies in econometrics Since a number of mathematizations in econometrics and in factor analysis are similar even though the substantive problems appear to be different, the last statement is not a real surprise A real surprise occurs when scholars in factor analysis and econometrics, for example, realize that their mathemat

(*) Table III reprinted from Louis Guttman "A New Approach to Factor Analysis the Radex" Chapter 6 in *Mathematical Thinking in the Social Sciences*, edited by Paul F Lazarsfeld, 1954, p 329

ical models are similar, for then the same mathematical developments can suffice for all.

If the circumplex is not equally spaced, the correlation matrix is no longer a circulant. However, there will still be a tendency for the largest correlations to be next to the main diagonal and in the upper right and lower left corners. Unequal spacing will occur in an additive uniform circumplex whenever any two $\sigma_{c_i}^2$ are unequal.

Guttman's discussion of modifications of the perfect simplex and perfect circumplex models and the relationships of these modifications to the Spearman model and common factor models is interesting but discursive. He gives, among others, an analysis in which two alternative extensions of a simplex, one the familiar additive model for common factors and the second a multiplicative system of common factors, both lead to the correlation matrix for the simplex structure. Naturally other functional forms might also lead to this structure. This indeterminacy has come up before in this paper and is essentially one of the targets for discussion.

Guttman combined his simplex and circumplex structures into the Radex concept. However, he is not so specific here as in his simplex and circumplex structures and attempts only a schematic description of the model. In the general radex model, a test response is a function of both complexity and kind of test and thus no longer belongs strictly to either a simplex or a circumplex. By this two dimensional mapping Guttman hopes to achieve a representation of mental abilities and a method for predicting them by the smallest number of available tests. One point, continually emphasized by Guttman, is the possibility of prediction with his factor analysis model, a trait not directly enjoyed by any of the other factor analysis models. In all other models, prediction techniques must be created in addition to and in accordance with the model employed. For example, if a Spearman model has been employed, there still remains the question of an appropriate function of the parameters of the model to be used for prediction or classification purposes.

Since Guttman's work is rather recent, it has naturally not received the attention given to the Spearman and Thomson models nor that given to the common factor theorists. The Radex is certainly feasible as a model and also serves in several instances as a possible explanation for data resulting from tests of mental ability. No doubt there will be additional discussion of this model in the literature.

8 THE COMMON FACTOR MODELS

We now return to the point from which we embarked on the Guttman studies and consider other alternative extensions of the Spearman model. We will first give an historical account of the development of common factor models and in later sections present the models of Holzinger, Thurstone, and Hotelling in some detail.

The common factor models began to make their appearance in the early nineteen thirties. They are natural extensions of the theory of General Ability. In fact, given the existence of the Spearman model they seem to be more natural extensions than the Guttman model.

It appears safe to say that the initial developments in the movement away from Spearman's single factor supposition were strict mathematical operations, rather than intuitive conjectures. Truman Kelley's [1928] first book on measurement of mental ability prepared in the mid nineteen twenties studied necessary and sufficient conditions for some mathematical theorems related to the question of number of tests and number of factors. Kelley's book contained a minor contribution by Harold Hotelling in connection with a problem propounded by Kelley when they were both at Stanford University. Interestingly enough, Hotelling was to make his fundamental contribution to factor analysis several years later through the efforts of Kelley, and Kelley was also destined to play a more prominent role in the development of the mathematics of multiple common factor models.

Motivated by the need for mathematical exploration of extensions of the Spearman type of analysis, the Carnegie Corporation in 1931 made a grant to E. L. Thorndike through the American Council on Education for an exploratory study of unitary differential traits in human nature. However, much more than mathematical theory and techniques were to be investigated in this study. For example, consideration was to be given to methods suggested by advances in physiological and comparative psychology under the direction of K. S. Lashley; a similar analysis was to be made in mental pathology by T. V. Moore; a similar investigation was to be made of common factors for an identical population by Holzinger and Spearman, and experimental studies of the analytic methods of Kelley, Spearman, and Thurstone were to be considered by Henry Garrett and Clark Hull. Kelley in cooperation with Hotelling was charged with the responsibility of supplying mathematical theory and techniques necessary for the development of factor analysis.

Hotelling's fundamental paper [1933] on the theory of principal com

ponents was the result of his work with the Unitary Traits Committee The method of principal components was essentially proposed in another connection in an early article by Karl Pearson [1901], but Hotelling's work gives a precise and more elegant treatment with direct reference to factor analysis and the sampling distribution theory related to the estimates of the factor loadings Hotelling in later papers also made substantial contributions to important computational problems in factor analysis Kelley's work was also essentially the same as Hotelling's development and went by the name of the method of principal axes

Shortly before Hotelling's fundamental contribution, Louis Thurstone was developing the mathematics of his procedure, called the centroid method, for multiple common factor models Thurstone's first efforts were mathematical and were related to factorial studies of mental abilities The application of these factorial methods in the isolation and identification of a number of primary abilities followed the efforts given to mathematical developments Several early specific applications of the method were the identification of primary factors in the vocational interests of high school graduates, and the identification of primary mental abilities from a set of sixty psychological tests given to 240 subjects A procedure by Cyril Burt, called the summation method, has much in common with the centroid method The centroid method is used almost exclusively by the American school of factor analysts except in some isolated instances where the method of principal components is still featured In addition, the psychological constraints promoted by Thurstone are invariably followed in this country while British investigators follow Burt's leadership in factor analysis studies

It may appear strange to interject the Bifactor model of Karl Holzinger at this point since by its structure it serves as a step between the Single Factor Theory and multiple common factor models However, historically, it reaches us after the Hotelling and Thurstone thinking on factor analysis The Holzinger model is an attempt to maintain the Spearman model by making some compromises on group factors but it does not yield on the sole common factor assumption Nonetheless it does give up the notion of all tetrad differences equal to zero

With the development of the multiple common factor models we have begun the departure from strict measurement of mental ability, at least in terms of parameters which could have some operational relationship to the physiological structure of human mental ability This, of course, does not mean that important and useful structures cannot be created for other purposes Moreover, the persistent existence of manifest correlation matrices

with ranks greater than one or two would require more than the models we have already considered even though the new models would no longer serve as strict measurement devices of mental ability In addition, the adventure of creation of new mathematical models as extensions of those already discussed could provide new structures which need not even reproduce existing data

There came a point where it became evident from many empirical examples that Spearman's original model had been superseded by a model containing many factors The Bifactor model could be looked upon as a step in this direction but it was still in the tradition of a lone general intellective factor Under these circumstances it comes as no surprise that several individuals turned their attention to multiple factor models and to methods by which any matrix of test correlations could be transformed directly into the latent multiple factor structure which allegedly produced it

As we shall soon observe, the multiple common factor structure is essentially the same for Thurstone and Hotelling, a major difference exists in the rationale for the estimation of the factor loadings However, in both the Thurstone and Hotelling developments, the data resulting from responses to tests is given the heavy burden of providing all the information on the factor loadings One can view the Spearman and Holzinger models from the structure of multiple common factors where a) in the Spearman model all factor loadings for the common factors except one are made equal to zero, and b) in the Holzinger model a similar vanishing of factor loadings except for more group factors, is achieved to yield the Bifactor structure The important issue, now, is why in the multiple common factor structure we have discussed can we not make several and sundry factor loadings vanish before we ever view the correlation matrix It could be possible just by expert judgment, conjecture, or theory, to give zero values to factor loadings especially since this is what Spearman and Holzinger did, or at least what they did can be viewed in this manner

Neither Thurstone nor Hotelling attempt this directly although in each approach, as we will see, some attention is given to this point albeit in a negative way Obviously one reason for as many zero factor loadings as possible is the resulting simplicity of the model Another indirect benefit is that the fewer parameters to be estimated, the more powerful and informative will be the estimation of each factor loading from a fixed set of manifest data Econometricians, who from other contexts end with the same linear structure as our multiple common factor structure, put in as many zero coefficients (factor loadings in the language of this essay) in their model

equations as they possibly can by expert judgment before they proceed with the estimation of the remaining coefficients Thurstone's rationale, as we shall see, leads to the statement that there are a number of zero factor loadings but their position is not specified

By now the reader can guess that both Thurstone and Hotelling demand that the data themselves produce the zero factor loadings although from two quite different approaches To give the Thurstone approach to zero factor loadings it is best to return to the very beginning of his work which brings us back to the point from which we strayed several paragraphs above Similarly, the question of zero factor loadings in Hotelling's procedure would best await the full development of his method of principal components However, despite a controversy which probably equaled the Spearman-Thomson battle in intensity if not in length of time, the Hotelling-Thurstone enthusiasts both employed the same mathematization, namely the multiple common factor structure The controversy was over the manner of estimating the factor loadings and the interpretations given to them as contrasted with the Spearman-Thomson debate over mathematizations leading to zero tetrad differences

9 THE HOLZINGER MODEL

One of the major points of controversy between Spearman and Thomson was on the relationship of group factors and the general factor in the latent models which produced hierarchical order or zero tetrad differences Spearman's thinking claimed only a general factor with the implicit statement that if correlation matrices were indicating group factors this manifestation could be appropriately handled by better testing procedures On the other hand, Thomson demonstrated that hierarchical order could be produced without the existence of a general factor and in the presence of overlapping group factors

Holzinger [1935], who worked in the Spearman tradition, provided a natural modification, as contrasted with extension, of the theory of General Ability using as a point of departure the place of group factors in the latent model By the time of Holzinger's formulation in the mid 1930's many empirical studies yielded tetrad-differences bounded away from zero Holzinger's model is an attempt to preserve the notion of a single general factor, even under non zero tetrad differences, by providing for *mutually exclusive* group factors without any additional statement such as

Spearman's that more appropriate testing techniques could remove the group factors That is, Holzinger's model assumed that a test score is a linear function of the sole general factor, a group factor stemming from the test under consideration (any test having a group factor cannot have another group factor), and a factor specific to the test only Spearman's model, of course, contained no group factors but did contain mutually exclusive specific factors One way, then, of looking at the Holzinger model is to yield to the notion that mutually exclusive group factors cannot be eased out of the model to present the pure Spearman general factor model Also mutually exclusive group factors might be viewed as an extension of mutually exclusive specific factors, that is, in the latter case an element is related to a test, in the former situation an element is related to a group of tests However, in the Holzinger model, a specific factor is attached to each test in addition to the single group factor and single general factor In brief, then, since zero tetrad-differences or correlation matrices of rank one were often not produced, Holzinger's suggestion is that there is an added component due to the group in which a test falls Notice now that we are beginning to get away from strict models of mental ability In fact, we now distinctly include in a model a parameter related to the construction of test batteries

Let us now look at the mathematization of this model If F is the lone general factor, G_i the group factor for the i^{th} group, and S_j the specific factor for the j^{th} test in the i^{th} group ($j = 1,2, \ldots, n$), then Holzinger's Bifactor model states

$$z_{j_i} = a_{j_i}F + a_{j_i}G_i + S_j \tag{9 1}$$

where z_{j_i} is the observed response to the j^{th} test in the i^{th} group and a_{j_i}, a_{ij_i} are the factor loadings or parameters of the model Once again we can assume the variances of the test responses are unity and evaluate the correlation matrix produced by this model The assumptions are similar to the Spearman model, namely, F, G_i and S_{ij} are independent over all i and j, variances of F and G_i are unity, variance of S_i is σ_j^2, and their means are zero Now first consider any two tests in the same group, say test j and test k in group i Then

$$Ez_{j_i}z_{k_i} = Ea_{j_i}a_{k_i}F^2 + Ea_{j_i}a_{k_i}G_i^2 + ES_{ij}S_k + \text{remaining terms} \tag{9 2}$$

All the remaining terms will be zero by the independence assumption, $ES_{ij}S_k$ equals zero for the same reason, and we get

$$r_{j_i k_i} = a_{j_i} a_{k_i} + a_{ij_i} a_{ik_i}. \tag{9.3}$$

This result is similar to Spearman's except for the added term which is due to the common group factor. For any two tests not in the same group, say j_i and d_c, we would get, following the same analysis,

$$r_{j_i d_c} = a_{j_i} a_{d_c} \tag{9.4}$$

or exactly the same result as Spearman since the overlap is now due solely to the general factor. On these bases, the total correlation matrix will now have rank two except for the empty diagonal cells but these can be uniquely filled to accomplish this rank.

A desire to check into the tenability of the Holzinger model brings up a type of problem not directly encountered in our previous models. Namely, the grouping of the tests must be established before we assess the model against data given in an empirical correlation matrix. This is not usually an easily resolved problem. Naturally, content of the tests is a helpful index but in addition grouping techniques applied to the data resulting from test responses must be used. We have mentioned before that one of the important uses of factor analysis is as a research technique for classification or grouping purposes, yet in this situation we are faced with a classification problem before we can apply the factor analysis. Once the grouping of the tests is accomplished, the reproducibility of the Holzinger model can be ascertained although one must always exercise caution in a decision of untenability because this can be a function of inappropriate grouping rather than the latent structure of the model itself.

10. THE THURSTONE MODEL

Louis Thurstone [1932] noticed that one solution to the factor analysis problem could be reached by a generalization of the Spearman idea of zero tetrad-differences or in other words by examining the rank of the correlation matrix. This and his other ideas on multiple factor analysis are given in two subsequent publications: an early (Thurstone [1932]) and a revised (Thurstone [1947]) edition of his fundamental book. Implicit in what we have said about the common factor generalization from the Spearman model is the assumption of a linear relationship between test response on the one hand and common factors and a specific factor on the other. Thus the rank of the correlation matrix, provided the diagonal elements do not interfere, is exact-

ly equal to the number of common factors In addition to the determinantal minors of order two, we can examine those of order three and all higher orders (still avoiding the diagonal cells, i e, the correlation of a test with itself) These determinantal minors can also be viewed as higher order tetrad differences For example, a three rowed determinantal minor (no diagonal cells) can be written as a function of a two rowed determinant in which each of the four elements is a tetrad difference We can then take a tetrad difference of these four tetrad differences If this second order tetrad difference is zero then the three rowed determinantal minor is zero If all three rowed determinantal minors are zero (that is, all second order tetrad differences zero) but one two rowed determinantal minor is not, then the rank of the correlation matrix is two and the tests can be imbedded into a structure with two common factors In other words, the test responses can be related to as many common factors as the rank of the correlation matrix, that is the rank apart from the diagonal cells, plus a specific factor for each test

For the mathematization of Thurstone s hypothesis we write

$$z_j = a_{j1} F_1 + a_{j2} F_2 + \quad + a_{jn} F_n + S_j \tag{10 1}$$

where z_j is the response of the j^{th} test, $j = 1, 2, \quad, n$, there are $m < n$ common factors F_1, S_j is the specific factor for test j, the a_j s are the factor loadings for the j^{th} test on each of the common factors While oblique factors can be analyzed, in general, we assume the common factors and all specific factors are mutually independent Also we usually assume each of the common factors is normally distributed with zero mean and unit variance S_j with zero mean and variance σ_j^2, and z_j with zero mean and variance equal to one Under these conditions, we can write the correlation r_{jk} between any two tests, j and k, as

$$Ez_j z_k = r_{jk} = a_{j1} a_{k1} + a_{j2} a_{k2} + \quad + a_{jm} a_{km} \tag{10 2}$$

This equation, which relates the manifest data of the correlation matrix to the latent parameters of the multiple common factor model is sometimes called the fundamental equation of factor analysis If there is only one common factor, note that we have Spearman's relationship For the Holzinger model we obtained the first term on the right hand side of the equation for pairs of tests not in the same group, and the first two terms for pairs of tests in the same group

Another way of arriving at the fundamental equation (10 2) is by the partial correlation argument used previously in our development of Spear

man's efforts. If we consider the partial correlation coefficient between any two tests, "partialing out" all the common factors, namely $r_{jk\cdot123\ldots m}$, the result is zero since there will no longer be any overlap. Setting the numerator of $r_{jk\cdot123\ldots m}$ equal to zero yields the fundamental equation (10.2).

Thurstone, of course, is saying that the number of common factors, namely m, should generally be small when compared to the number of tests in a battery. If parsimony does not exist, then the desire to factor can easily be diminished unless one is especially interested in achieving n orthogonal dimensions for the observed n correlated dimensions. Ordinarily a regular correlation matrix, based on n tests, with any elements in the main diagonal would have a rank equal to n. Up to this point we have been quite cavalier about the elements in the main diagonal of the observed correlation matrix, essentially saying that the vacant elements can be chosen appropriately and with due regard to the model under consideration. For example, in the Spearman model it was always possible to fill the vacancies with unique values to make the rank of the correlation matrix equal to unity. In multiple common factor models we depend on the manifest data to determine dimensionality of the common factor space or equivalently the rank of the correlation matrix. Thus while in a specific context the rank may be assumed unique and much less than n, both the rank and the "true" elements in the main diagonal, namely the communalities, have to be obtained by iterative methods. If the question is uniquely resolved, then both rank and communalities are determined at the same time; that is, neither is found first to yield the other. A diagonal element, now called a communality, can be looked upon as the contribution to the total variance in a test response due to the factor loadings of the common factors and thus of course need not be equal to one if we tolerate specific factors.

Suppose we now place ourselves in the position of assuming the multiple common factor structure of Thurstone as the latent model for an observed correlation matrix. The first task is the joint resolution of the dimensionality of the common factor space and the estimation of the factor loadings. A guess must be attempted for the presently vacant elements. As long as the estimate is not less than the true communality, the centroid procedure for estimating the factor loadings will prevail although some iterations of the procedure may be necessary to get the exact communalities. If the estimated diagonal cells are smaller than true communalities, the elements of the correlation matrix no longer form a positive definite quadratic form whose existence is essential for both the Thurstone and Hotelling procedures. Obviously one

way out of the dilemma is to put one's in the main diagonal even though this may be at the expense of economy of computations

Once the diagonal cells are validly chosen, the factor loadings are obtained by the centroid method The rationale behind this procedure is similar, as we shall see, to Hotelling's argument Namely, we choose the factor loadings, say for the first factor, so that their contribution to the total variation in the system (sum of variances of each test in the battery) is relatively maximized One quasi-analytic way of accomplishing this is to begin by making the first factor go through the centroid of all the test scores viewed in the m-dimensional common factor space (the point in common factor space whose m coordinates are the m arithmetic means each based on n factor loadings) This will tend to maximize the contribution of the first "centroid" factor to the total test variance but of course will not provide the unique maximum However, it does provide for much less computation than achieving the maximum Details of the centroid procedure can be found in many texts Once the loadings for the first factor are obtained, the effect of the first factor is removed from the correlation matrix A direct approach leads to an impasse in obtaining the factor loadings for the second factor because the centroid is now at the origin However, Thurstone proposed a method of "reflection" which avoids this obstacle Reflection is a mathematical operation and bears no relationship to the principles of Thurstone's model After the second factor is extracted, the procedure is continued until the elements in a residual correlation matrix are approximately zero It is, of course, hoped that this will occur when m is much smaller than n At this point, the dimensionality of the common factor space is evident (i e, the rank of the correlation matrix is determined) and the factor loading are determined

Also, at this point, comes the major point of controversy between the Thurstone and Hotelling models Now that dimensionality of common factor space has been established, Thurstone feels that the "centroid" factors are but one of an infinite number of settings of the m orthogonal axes (factors) that can be obtained by rotation of the orthogonal axes obtained by the centroid method from the observed correlation matrix Thurstone now demands that attention be given only to the one setting that achieves psychological relevance and he defines this as the one that achieves "simple structure" By simple structure we coordinate the belief that underlying causes in nature are fundamentally plain and simple with the notion that responses to a single test should not depend on all common factors but probably on a small subset of them Thus, in our rotations of the original factor loadings we seek the

setting that has as many zeros (or elements approximately zero) as possible This is of course made more specific by Thurstone by suggesting how many zeros should appear in a row or column of the factor loading matrix after an appropriate number of rotations Coupled with these statements is the notion that simple structure is an invariant and that a new test battery will provide a correlation matrix which after factoring and rotation will lead to the same simple stucture of factor loadings We have already raised the question of introducing simple structure before factoring and thus allowing the data on hand to provide more efficient estimates of the non-zero and thus unknown factor loadings But we do not reintroduce it here as a criticism of Thurstone's rationale but simply as another alternative approach to the problem especially since the econometricians have considered it

Suppose we now have a completed factor loading picture obtained by Thurstone's procedures on a multiple common factor model Apparently we have given up a direct reconciliation with measurement of mental ability related to the human brain and nervous system but it is not too difficult to see that valuable information about the construction of mental tests can result and that in this way measurement of mental ability is indirectly served

11 THE HOTELLING MODEL

Much of what we have said in the previous sections on the structure of multiple common factor models needs no repeating for consideration of the Hotelling procedure The structure is formally the same except that now we consider *n* tests and *n* factors and do not necessarily distinguish, at least at first, between common and specific factors The diagonal elements of the correlation matrix need not be unity but simply the true communalities to preserve the positive definite characteristic of the correlation matrix

The principal components procedure of Hotelling to estimate the factor loadings in the multiple common factor structure is a strictly mathematical operation designed to accomplish a specific maximization criterion very similar to the previously discussed centroid procedure Naturally it produces unique answers to a clearly defined and delineated mathematical problem but the interpretation of these answers within the context of our paper will bear some exploration

Once again consider

$$z_j = a_{j1}F_1 + a_{j2}F_2 + \quad + a_{jn}F_n, \quad j = 1, 2, \quad , n \tag{11 1}$$

and make the usual assumptions as before For example, the variance in each test is equal to one and the variation in the whole system is then n, the number of tests On this basis we can also write

$$\sigma^2_{z_j} = a^2_{j1} + a^2_{j2} + \quad + a^2_{jn} = 1, \quad j = 1, 2, \quad , n \tag{11 2}$$

However, as before, the fundamental equation of factor analysis applies This leads to $\frac{1}{2}\, n(n + 1)$ equations between the correlation coefficients and the factor loadings But there are n^2 factor loadings to be determined and thus the number of equations is insufficient for the task Thus systems of uncorrelated components (orthogonal latent factors) can be chosen in an infinite variety of ways and still reproduce the correlation matrix We have already discussed Thurstone's escape from this indeterminacy

Hotelling's specific proposal to achieve uniqueness of factor loading a_{ij} is as follows Subject to (11 2), namely, the condition that the sum of the squares across each row is unity, and the fact that the sum of the cross products reproduces the observed elements in the correlation matrix, we desire to maximize the contribution that each of the obtained factors can make to the total variance, n, in the system That is, we wish to make the sum of the squares of the obtained factor loadings in each column a maximum subject to the constraints we have placed on the system The sum of the squares of the factor loadings in a column is the contribution of that factor to the total variation Hotelling demonstrated that the problem of finding a component (factor) which will account for as large as possible a part of the total variance is solved by finding the largest root of the characteristic equation of a matrix Moreover, the ratio of this root to n gives the percentage of the total variation accounted for by the factor Interestingly enough from a mathematical model standpoint, the characteristic equations of a matrix studied by Hotelling in this context were mentioned by him as equations which were first studied in connection with the pertubations of the planets When the coefficients of the components (factor loadings) have been determined, the effect of this first factor is removed from the correlation matrix The next problem is to find the component (factor) making a maximum contribution to the residual portion of the variance The same procedure is applied until all factor loadings are obtained or equivalently the residual correlation matrix has only zero elements

Geometrically the foregoing procedure corresponds to rotating the n rectangular axes $z_1, z_2, \quad , z_n$ so that the new coordinate axes lie along the

principal axes of the ellipsoids defined by $\sum_{i=1}^{n} \sum_{j=1}^{n} r^{ij} z_i z_j =$ constant where r^{ij} is the element in the i^{th} row and j^{th} column of the inverse of the correlation matrix In fact, the squares of the lengths of the principal axes of any such ellipsoid are proportional to the roots of the matrix characteristic equations we mentioned previously This is why the Hotelling procedure for estimating factor loadings is called the principal components solution

At first glance the Hotelling procedure seemingly ignores parsimony since it computes all n^2 factor loadings However, as each factor is extracted we have knowledge of the percentage of variability remaining in the system One can decide to stop factoring after 95 per cent or some other large fixed percentage of the variation has been extracted and thus end with many fewer factors than tests The variation that remains can be ascribed to specific factors or just simply to sampling fluctuation In fact, the Hotelling procedure, since it is a specific probabilistic procedure, permits the construction of tests of significance for testing that factor loadings are zero although the actual development of these tests may be intractable in mathematical closed form Tests of statistical significance for factor loadings estimated by Thurstone's procedure are out of the question because of the quasi-analytic nature of Thurstone's development

Whatever the final outcome of the method of principal components, Hotelling essentially regards it as the resolution of the factoring problem There are no thoughts of rotation to achieve simple structure and no consideration to operate on the principal components resolution to achieve any other criteria meeting psychological relevance Of course, there is no reason not to rotate to simple structure once the Hotelling resolution is on hand But it does not seem prudent to undergo the more extensive calculations of the principal components method if one is interested in Thurstone's simple structure Perhaps this is why most texts do not explain or even mention Hotelling's method but are written entirely around Thurstone's centroid method and subsequent rotations to simple structure

12 SUMMARY

We have now discussed the major mathematical models in factor analysis produced by scholars over the past half century Perhaps the development of these models will be the only major contribution made in the math-

ematical facet of factor analysis Much work, at present, on sampling and distribution theory of estimates of factor loadings, not discussed in this paper, may be indicative of the sterility of effort from the viewpoint of measuring mental ability or of providing guides in the construction of tests of mental achievement On the other hand, new models may be in the offing as attempts are made to describe test responses as the result of a sequence of time changes in the brain and nervous system With the introduction of dynamic models, new developments in the mathematics of stochastic processes could have an impact on the development of measurement devices for mental ability To date, certainly, all the mathematizations have been of the static (non time parameter) variety

Much has been made in this essay of the notion of reproducibility of data The major models discussed have all enjoyed this feature to some extent Certainly any model which cannot re create an observed situation should be immediately suspect no matter how logical or persuasive the arguments that produced it The question of which of several tenable models is valid is always an interesting one but the resolution of this question is not necessary for tenable models to be both instructive and useful For example the multiple common factor structures may be a departure away from neurological measurement of mental ability but because they can usually replace a set of correlated dimensions with a smaller number of orthogonal axes, they have a use wherever such a replacement can present a situation in a clearer light Thus there have been many applications of the multiple common factor structure, as an exploratory probing tool to other areas of psychology and other scientific disciplines It is in this context and usually only in this context, that many students of factor analysis view the subject at present

BIBLIOGRAPHY

Anderson, Theodore W. and Rubin, Herman: "Statistical inference in factor analysis," *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 5, p. 111—50, University of California Press, 1956, Berkeley and Los Angeles

Galton, Francis: "Co-relations and their measurement, chiefly from anthropometric data," *Proceedings of the Royal Society*, Vol. 45, 1888, p 135—40.

Guttman, Louis: "A new approach to factor analysis: the Radex," Chapter 6, *Mathematical Thinking in The Social Sciences*, edited by Paul F. Lazarsfeld, Free Press, Inc., 1954, Glencoe, Illinois

Holzinger, Karl: *Introduction to Bifactor Theory*, Preliminary Reports on Spearman-Holzinger Unitary Trait Study, No 5, University of Chicago Press, 1935, Chicago.

Hotelling, Harold: "Analysis of a complex of statistical variables into principal components," *Journal of Educational Psychology*, Vol. 24, 1933, p 417—41 and 498—520.

Kelley, Truman L. *Crossroads in the Mind of Man*, Stanford University Press, 1928, Stanford, California.

Mackie, J: "Mathematical consequences of certain theories of mental ability," *Proceedings of the Royal Society of Edinburgh*, Vol. 49, 1929, p. 16—37.

Pearson, Karl. "On lines and planes of closest fit to systems of points in space," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol 2 (6th series) 1901, p. 559—72.

Solomon, Herbert and Rosner, Benjamin: "Factor analysis," Chapter V, *Review of Educational Research*, December 1955, p. 421—36.

Spearman, Charles: "General intelligence objectively determined and measured," *American Journal of Psychology*, Vol. 15, 1904, p. 201—93.

Thomson, Godfrey H.: "A hierarchy without a general factor," *British Journal of Psychology*, Vol. 8, 1916, p. 271—81.

—: "The hierarchy of abilities," *British Journal of Psychology*, Vol. 9, 1919, p. 337—44.

—: "The tetrad-difference criterion," *British Journal of Psychology*, Vol. 17, 1927a, p. 235—55.

—: "A worked-out example of the possible linkages of four correlated variables on the Sampling Theory," *British Journal of Psychology*, Vol. 18, 1927b, p. 68—76.

—: "On complete families of correlation coefficients, and their tendency to zero tetrad-differences: including a statement of the Sampling Theory of Abilities," *British Journal of Psychology*, Vol. 26, 1935, p. 63—92.

Thurstone, Louis L.: *The Theory of Multiple Factors*, University of Chicago Press, 1932, Chicago.

—: *The Vectors of Mind*, University of Chicago Press, 1935, Chicago.

—: *Multiple Factor Analysis*, University of Chicago Press, 1947, Chicago.

Wolfle, Dael: *Factor Analysis to 1940*, Psychometric Monograph No. 3, University of Chicago Press, 1940, Chicago.